MicroRNA Target Prediction Based upon Metastable RNA Conformations

Abdelhadi Ep Souki, Ouala

Awarding institution:
King's College London

MicroRNA Target Prediction Based upon Metastable RNA Conformations

Ouala Abdelhadi Ep Souki

Department of Informatics

King's College London

A thesis submitted for the degree of

Doctor of Philosophy

London, 2016

Acknowledgements

First and foremost, I consider myself truly privileged to work under the supervision of Dr. Kathleen Steinhöfel and Prof. Andreas A. Albrecht from Middlesex University London. Their extensive knowledge enhanced my understanding on my area of study. I am grateful for their valuable feedback on my works, and their constant guidance, advice and encouragement throughout the challenging period of my PhD study. I gratefully acknowledge the funding received towards my PhD from the EPSRC Doctoral Training Grant. Words cannot express how grateful I am to my parents, Samia and Khaled, and my parents in law, Saloua and Ali, for all their love and encouragement throughout this truly life-changing experience. Last but not the least, I would like to dedicate this thesis to two very special persons, my husband Imed and my son Adam, for their continued and unfailing love, emotional support and understanding during my PhD.

Abstract

MicroRNAs (miRNAs) play an important role in biomarker research. Identifying their targets and inferring their functions have been of a great importance to developing our understanding of many biological processes and fundamental novel anti-cancer and viral therapies. Since the discovery and validation of true miRNA-messengerRNA (mRNA) bindings is a laborious and expensive process, computational tools for the prediction of miRNA targets are essential in this research area. Advanced tools of miRNA target prediction incorporate knowledge about secondary structures of mRNA sequences, usually the 3'UTR into the evaluation and assessment of putative miRNA-mRNA bindings. The default secondary RNA structure in most target prediction tools of this type is the minimum free energy conformation or a representative of the ensemble of all possible RNA structures. A key indicator of putative miRNA-mRNA bindings is the energy required to open base pairs that are present in the potential binding site within the conformation. However, mRNAs as well as miRNAs are present in a single cell in multiple copies, where the number of copies may range from several tens up to several hundreds of copies, each of them transcribed from DNA at different points of time and therefore, potentially, being present in different folding stages, most likely in metastable conformations. In this thesis we have addressed the problem of miRNA bindings to metastable RNA secondary structures in the context of Single Nucleotide Polymorphisms (SNPs). To this end, we first searched the recent literature for disease-related triples [mRNA/3'UTR; SNP; miRNA] that have been analysed by methods including PCR and/or luciferase reporter assays. We next compared results of two major computational approaches to miRNA

target ranking prediction: conservation feature using `TargetScan` tool and target site accessibility feature using `PITA` and `STarMir` tools. We showed that site accessibility may be a better ranking criterion. We then studied the problem of miRNA bindings to metastable secondary structures in the context of SNPs and mRNA concentration levels i.e. whether features of miRNA bindings to metastable conformations could provide additional information supporting the differences in expression levels of the two sequences defined by a SNP. We showed that among the different parameters we introduced and analyzed, we found that three of them, related to the average depth and average opening energy of metastable conformations, may provide supporting information for a stronger separation between miRNA bindings to the two alleles defined by a given SNP. These findings were a trigger to devise a novel target prediction tool that incorporates metastable secondary structures with low energy levels into predictions. We present, `RNAStrucTar`, a miRNA target prediction tool that analyses putative mRNA binding sites within 3'UTR secondary structures representing metastable conformations. The first stage consists of generating conformations that can be classified as deep local minima. The second stage incorporates duplex structure prediction through sequence alignment and energy computation. Target site accessibility related to different sets of metastable conformations is also taken into account. An overall interaction score computed from multiple binding sites is returned. The approach is discussed in the context of SNPs where our manually curated [mRNA;SNP;miRNA] dataset is utilised. `RNAStrucTar` predictions are in favour of the allele with the stronger miRNA binding stated in the underlying literature in 22 instances, while the resulting scores are indifferent in ten cases. For the two other cases (HTR3E and FGF20), the score is in favour of the weaker allele. In this respect, `RNAStrucTar` results are better than `PITA` and `STarMir`, with a positive prediction for RAD51 and MSLN (`STarMir` favours the weaker allele).

Contents

Chapter 1

Introduction

In February 2001, the Human Genome Project (HGP) published the first draft of human genome [1]. Since then, many other whole-genome sequencing projects have been completed which have elucidated the complexity of the human genome and subsequently the human proteome. Building on the work of the HGP, the Encyclopedia Of DNA Elements (ENCODE) project was carried out to catalog and describe all the functional elements in the human genome [2]. The completion of other projects, such as the SNP Consortium [3] and the HapMap Project [4], has also impacted the field of RNA research in profound ways. In this chapter, we first introduce the biological aspects of RNA folding and then we discuss the microRNA target prediction problem by explaining the associated computational principles and reviewing state of the art prediction tools. We conclude the chapter with an overall organisation of the thesis.

1.1 RNA Folding Concepts

1.1.1 RNA Biogenesis

The central dogma of molecular biology explains the flow of genetic information, from DNA to RNA, to make a functional product, a protein. The nucleic acids, i.e. DNA (Deoxyribonucleic Acid) and RNA (Ribonucleic Acid) serve as storage for genetic information. RNA and DNA are polymers of repeating units of monomers called nucleotides, made from three distinct modules that are a sin-

gle sugar, a single base and up to three phosphate groups. Each nucleotide has one of four types of bases, namely Adenine (A), Uridine (U) (Thymine (T) in the case of DNA), Guanine (G) and Cytosine(C). Several forms of RNAs include messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). mRNAs are short-lived protein-coding RNAs that carry information from DNA to the ribosome sites to generate protein.

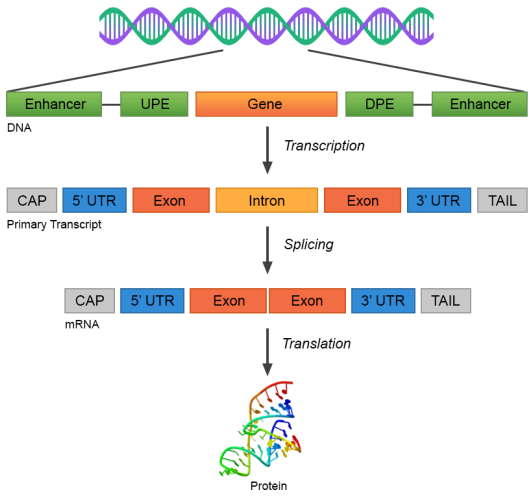


Figure 1.1: The central dogma of molecular biology.

Figure 1.1 illustrates the different biological steps from DNA to protein. The transcription, splicing and translation are the main processes that account for gene expression of protein coding genes. Transcription is the copying of DNA (template strand) to RNA (pre-RNA). Pre-RNA consists of two types of sub-sequences: exons and introns. Introns are sub-sequences between exons that are removed by splicing; the remaining exons form the RNA sequence, where one distinguishes between two main types: messenger RNA (mRNA) that carries information about proteins (encodes protein information), and non-coding RNA. Messenger RNA is transformed into protein by a process called translation (within the coding region of mRNA, three consecutive nucleotides - called codons - define a single amino acid in the protein). After splicing, the mRNA consists of the central coding region (CDS) and a flanking region (in most cases) on either side of the coding region, which are called untranslated regions (UTRs). Thus, the untranslated regions are sections of the mRNA located before the start codon

and after the stop codon of the coding region, termed the five prime untranslated region (5' UTR) and three prime untranslated region (3' UTR), respectively. Figure 1.2 illustrates a typical mature mRNA. The mRNA also contains two other special regions: the 5' cap added to the 5' end which provides stability during translation and the 3' poly (A) tail added to the 3'end which protects the mRNA from degradation.

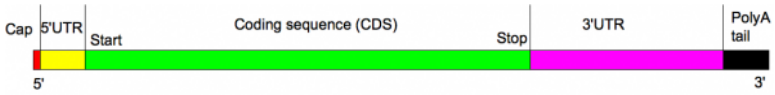


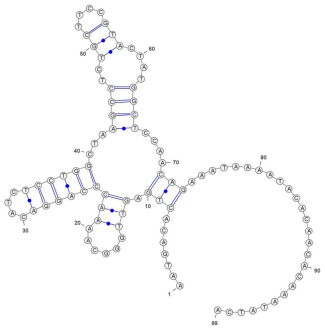
Figure 1.2: The structure of a typical mature Messenger RNA.

1.1.2 RNA Folded Structure

An RNA molecule can be described as a hierarchical structure in which the primary sequence determines the secondary structure which, in turn, determines its tertiary folding, whose formation alters only minimally the secondary structure [5]. The secondary structure is the folded form of the nucleotide chain that is a 2-dimensional structure where certain substructures such as hairpins or loops are formed through base pairing. It provides an ideal model to study RNA folding, provides a biochemically useful structure description, mathematically and computationally easy to handle, and uses energy models based on carefully measured parameters. RNA molecules are also able to fold into tertiary shapes which can bind to other molecules very tightly and specifically. It is this 3-dimensional conformation that determines RNA function. Figure 1.3 illustrates the different structural levels of an RNA sequence, where the secondary structure drawing was produced by `Varna` visualization tool [6] and the tertiary structure generated by `RNAcomposer` [7].

A secondary structure consists of a number of substructures or motifs e.g. single-stranded regions, stacked base pairs, hairpin loops, multiple loops, interior loops, and bulge loops (see Figure 1.4). Hairpins (or stemloops) are the most common element of RNA secondary structure [8; 9]. The foundation of the RNA structure formation is a continuous base pairing, resulting in helical regions or stems. A helical stem is also referred to as a stacking region. A hairpin loop is

AATGACACTGAGT
TGGGCAAAACCCA
GGACATCTCCTGG
CTAAGCCTCTGCT
TCCGTACTATGGC
TCCAACAGAAATA
AAATACACAACAC
AAATATCA



(a) Primary sequence (b) Secondary structure (c) Tertiary structure

Figure 1.3: The RNA structure hierarchy.

a single stranded region which folds back on itself. When the parts of the RNA sequence spanned by two base pairs are neither disjoint, nor have one contained in the other, the two base pairs form a pseudoknot. A bulge loop occurs when a duplex is interrupted by one or more unpaired nucleotides in one of the strands. The objective of RNA folding in bioinformatics is to accurately model the structure of the RNA from its given sequence, where the most likely RNA structure to form is predicted by computational programs.

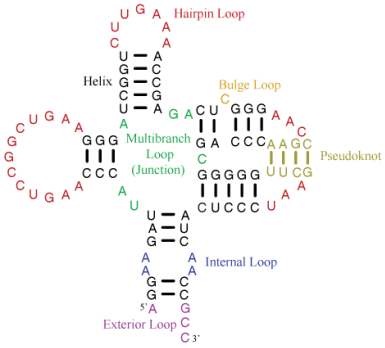


Figure 1.4: RNA secondary structure motifs (adopted from the Nearest Neighbor Database rna.urmc.rochester.edu).

1.1.3 Problem Formulation

We give in this section the well-known mathematical formulation of the secondary structure problem. The primary sequence of RNA is represented as a string s of n characters, $R = r_1, r_2, \ldots, r_n$ where $r_i \in \{A, C, G, U\}$. The secondary structure of a sequence R is the set of base pairs (r_i, r_j) such that the following properties are satisfied:

- For each pair is $(i, j), 1 \leq i < j \leq n$ i.e. each base of a pair must be from different locations.
- Each pair (i, j) is either a Watson-Crick pair $(r_i, r_j) \in$ $(A, U), (U, A), (C, G), (G, C)\}$, or a wobble pair $(r_i, r_j) \in \{(G, U), (U, G)\}$.
- Each index $k = 1, \ldots, n$ can only occur at most once in a pair.
- Sharp folds in the sequence are prohibited, that is $j - i > p$, where p is a positive constant. For example, a hairpin loop should have at least three unpaired nucleotides [10].

A pseudoknot occurs when two pairs, between nucleotides i and j and between nucleotides i' and j', exist with $i < i' < j < j'$.

Several representations of the secondary structure have been utilized. While the graph representation shown in Figure 1.3 gives an idea of the shape of the RNA sequence, the string notation known as the dot bracket notation is a machine readable representation. It is defined by the following rules:

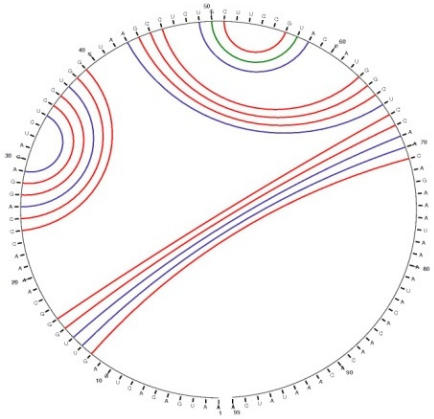
- If a base i is unpaired, then it is represented as $S_i =' .'$ in the string
- If $(p, q)$is a base pair and $p < q$, then $S_p =' ('$ and $S_q =')'$.

The first rule applies to all loop and bulge substructures. The second rule applies to helix or stacked stem substructures. For example, the string ((((.....)))) corresponds to a stacked helix of length 4 and a hairpin loop of length 5.

Figure 1.5 shows the dot bracket and the circular representations where nucleotides of a secondary structure are represented by a circle and arcs are drawn for each base pairing. If any arcs cross, a pseudoknot is present. The dot bracket

.....(((.(((......)))((((((....)))))))...((((.. (((....)))....))))
....)))...........................

(a) Dot Bracket representation of MFE structure



(b) Circular representation

Figure 1.5: Different RNA representations of the RNA sequence and secondary structure from Figure 1.3.

representation was produced by `RNAfold` [11] and the circular representation generated by `mfold` [12].

1.1.4 Secondary Structure Prediction

Research on RNA secondary structure prediction dates back to around 1960 [13]. Further progress was published in [14; 15; 16]. The method by Tumanyan *et al.* [16], published in 1966, comprises already the essential ingredients of a dynamic programming approach for RNA secondary structure prediction [17]. The algorithm was implemented with a potential of handling sequences of length up to n = 600 with the then available hardware. A more formal application of dynamic programming to RNA secondary structure prediction was carried out in 1978 by Nussinov *et al.* [18] and Waterman and Smith [19]. Nussinov algorithm compares a sequence against itself and finds the maximum of the scores for the possible structures at a particular position. Base pair maximization will not necessarily generate the most stable structure and may have scattered matches which are

not biologically reasonable. It may create a structure with many interior loops or hairpins which are energetically unfavorable. Therefore this algorithm does not give accurate structure predictions [20; 21].

Later on, the application of dynamic programming to secondary structure prediction by energy minimization, with various refinements and energy functions, eventually resulted in the design of powerful tools for folding simulations and associated program packages, such as `RNAfold` by Hofacker and co-workers [11; 22] and `mfold` by Zuker and co-workers [20; 23; 24].

Different secondary structures could be shaped from the same RNA sequence. Each structure has its specific free energy change $\triangle G$ introduced as the Gibbs free energy change of folding and expressed in KiloJoules/Mole (Kj/mol). These algorithms are used to search for a single secondary structure with the lowest free energy change named the minimum free energy(MFE) conformation. The Turner nearest neighbor energy model, where free energies are assigned to loops rather than to base pairs, using loop dependant energy rules, is widely used for free energy calculation [25; 26]. The key idea of this algorithm is that a secondary structure can be decomposed into a number of separate loop structures such that the total free energy of the structure is the sum of these parts. As an example, stacking pairs are the dominant stabilizing force as they contribute to the negative free energy while unpaired bases, such as interior loops, hairpin loops, bulges and multi-loops, form destabilizing loops, thus contributing with a positive free energy.

There is a great diversity in recent research on secondary structure predictions, including attempts to tackle the problem of folding simulations for structures with pseudoknots, which is known to be NP complete [27]. Kinetic folding simulations and modeling of co-transcriptional folding are an important consideration. Furthermore, sampling techniques focusing on approximations of the partition function over all secondary structures or specifically for metastable conformations have also been of major interest.

1.1.5 Kinetic folding

Kinetic folding simulations provide valuable information about transition rates at different folding stages and the stability of substructures such as helices, hairpins, various types of loops, and pseudoknots. Flamm and Hofacker provide an overview of methods for kinetic folding simulations in [28], see also the detailed summary by Schuster [29].

While basic kinetic moves are addition and deletion of single base pairs, Flamm *et al.* [30] introduced the shift move, which is a combination of a base pair removal and a base pair addition where one position remains invariant. The shift move aims at the simulation of 'defect diffusion' reported in [31], which tries to capture the process where the position of a bulge in a helix may move along a helix as the result of rapid base pair formation and dissociation. Geis *et al.* present the `Kinwalker` tool in [32], along with simulation results on twelve RNA sequences with a length ranging from 115 nt to 1492 nt. The underlying method generates secondary structures by combining building blocks representing thermodynamically optimal substructures. The optimal substructures are calculated by standard dynamic programming. Finding the energy barrier between two locally optimal conformations is an important step in the execution of `Kinwalker`. The sub-routine employs a modification of the Morgan-Higgs heuristic [33], which is designed to find a direct folding pathway of minimum height between two secondary structures. Most of the total run-time of `Kinwalker` is spent on this particular task. The authors report a good agreement of folding simulations with experimentally verified folding pathways, and the same applies to calculated folding times and corresponding values predicted in the literature.

Co-transcriptional folding is generally acknowledged as describing the process of how RNA folding happens *in vivo* [34]. As pointed out in [28] and [35], RNA is transcribed at a rate of only $\approx$ 30-40 nucleotides per second, where the nascent chain starts folding as soon as it leaves the ribosome. Since helices formed by the incomplete chain may be too stable to refold later on, co-transcriptional folding may drive the folding process to a well-defined folded state that is different from a MFE conformation. In a recent experimental study, Solomatin *et al.* [36] argue in favor of multiple RNA folding pathways to different biologically

active conformations (where the authors include the wider perspective of protein folding). Therefore, metastable structures have become the subject of recent research due to the new insights into co-transcriptional folding and interactions between different types of RNA sequences.

Studies on RNA folding kinetics may provide interesting insights into the kinetic mechanism of RNA functions. However, analyzing the underlying folding energy landscape, which will be defined in the following section, presents a simpler alternative.

1.1.6 Energy Landscapes

We briefly recall the main features of energy landscapes induced by RNA secondary structures. Given an RNA sequence R, we denote by $\mathsf{L}(R) = [\mathsf{C}, \mathsf{N}, E]$ the energy landscape defined by the set of secondary structures C, the neighborhood relation N and the energy function $E : \mathsf{C} \to \mathbb{R}$. The conformation space C consists of secondary structures with standard settings as provided, for example, by the `RNAfold` tool [11], i.e. no isolated base pairs and at least three nucleotides in loops.

Given a secondary structure S of sequence R, the neighborhood N_S is defined by two types of single-step transitions $S \to S' \in \mathsf{N}_S$:

(1) Addition of one or two base pairs: a single base pair is added, if an existing helix is extended; two base pairs are added, if an unpaired position admits such an extension without extending a helix by two base pairs; the addition must ensure that the condition for the minimum loop size is not violated.

(2) Deletion of one or two base pairs: a single base pair is deleted as part of a helix, if at least two adjoined base pairs remain; otherwise, two base pairs are deleted.

The neighborhood N_S covers all conformations that can be generated by a single application of one of the transitions, where by definition the secondary structure S itself belongs to N_S.

The `RNAsubopt` tool by Wuchty *et al.* [37] is an algorithm that generates all suboptimal folds of a sequence in a partial energy landscape, which is within a

desired energy range from the MFE. The idea underlying the algorithm is taken from Waterman and Byers [38] who developed it in the context of suboptimal solutions to the shortest path problem in networks. A more detailed investigation of the energy landscape of RNA is possible and defines the following particular features:

- A local minimum is a structure with an energy lower than the energy of all neighboring structures.
- A local maximum is a structure with an energy higher than the energies of all legal neighboring structures.
- A saddle point is a structure from where at least two local minima can be reached by a downhill walk starting with this structure.

Methods to elucidate the basin structure of landscapes by means of trees that represent local minima and their connecting saddle points have been developed. This includes the `Barriers` program [39], a tool that inputs an energy sorted list of conformations of a landscape, and computes local minima and energy barriers of the landscape. An example of a barrier tree is given in the Figure 1.6. Lorenz and Clote introduce in [40] the `RNAlocopt` tool for sampling and approximating the total number of metastable conformations (local minima) using the partition function. However, currently the `RNAlocopt` tool has only been implemented using the older Turner 1999 energy model without dangling ends. RNA energy landscape analysis in the context of metastable conformations has also been presented in [41] ,[42] ,[43] and [44].

1.2 MicroRNA Regulatory Targets

One of the most significant recent advances in Cell Biology is the discovery of small non coding RNAs, known to possess regulatory functions at the translational level in cells of various species. The term non-coding RNA (ncRNA) is commonly used for RNA that is not translated into a protein. It has been assumed that genes generally code for proteins [45]. However, recent evidence

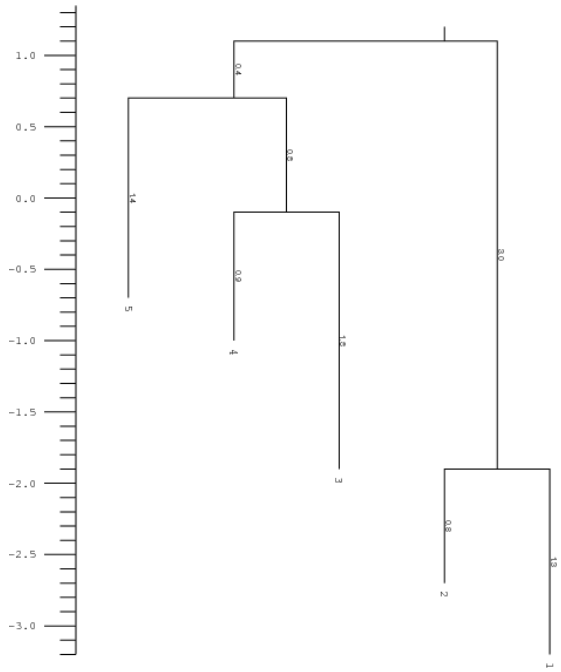


Figure 1.6: An example of a `Barriers` output tree where local minima are labeled with numbers and the height of the energy barrier to their connecting saddle points is given. The global minimum is marked as local minimum 1.

suggests that the majority of the genomes in mammals and other complex organisms is in fact transcribed into ncRNAs, many of which are alternatively spliced and/or processed into smaller products [46]. ncRNAs are involved in a varied number of cellular processes. Among other roles, they regulate gene expression at the levels of transcription, RNA processing, and translation. ncRNAs are divided into two main groups: the short ncRNAs (< 30 nts) and the long ncRNAs (> 200 nts). The three major classes of short non-coding RNAs are microRNAs (miRNAs), short interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). A detailed review of classes of ncRNAs and their functions is given in [47].

1.2.1 MicroRNA Biogenesis

MicroRNAs are short, 17 to 24 nucleotides long, non-coding RNAs encoded by a significant proportion of genes [48]. They are produced from either their own genes or from introns and rarely from exons [49]. The first report of a miRNA, lin-4, that regulates the development of *Caenorhabditis elegans* (*C. elegans*), dates back to 1993 [50; 51]. However, it was not until later, with the discovery of the let-7 miRNA and its conservation from worms to humans, that the functional importance of miRNA-dependant gene regulation emerged [52; 53]. In the following years, hundreds of miRNAs were identified in a wide range of species.

The current molecular framework for the mammalian miRNA biogenesis is shown in Figure 1.7. MicroRNA genes are transcribed by either RNA polymerase II or RNA polymerase III into long primary miRNA transcripts (pri-miRNA) [54; 55]. The precursor (pre-miRNAs) of $\approx$ 65 nucleotides in length [48] with a stem loop structure is next released from pri-miRNA by a cleavage event, which is catalysed by the nuclear microprocessor complex Drosha-DGCR8 (Pasha) in the nucleus [56]. The resulting precursor hairpin, the pre-miRNA, is then processed by another RNase III enzyme, Dicer and this causes the release of a double-stranded RNA duplex of $\approx$ 22 base-pair RNA that is composed of the eventual mature miRNA, base-paired to a complementary miRNA* strand, or the passenger strand [57]. Subsequently, one of the strands, designated as miRNA or the guide strand, is preferentially selected for maturation whereas the

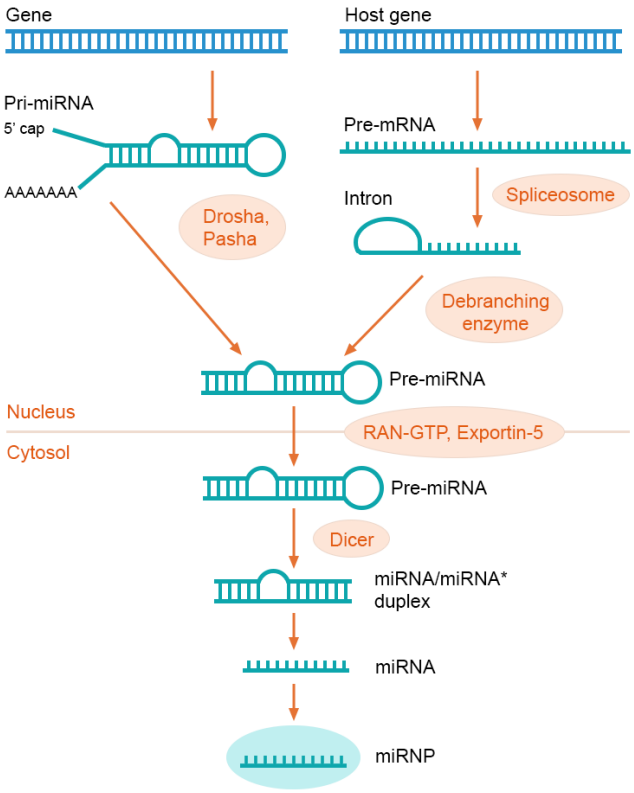


Figure 1.7: MicroRNA biogenesis.

passenger strand is degraded [58]. However, recent reports have indicated that in some cases, the two strands often co-exist and are both functional. In these cases, the mature miRNA species may be derived from both the 5' and 3' arms of the precursor duplex, and are called the miRNA-5p and -3p species, respectively [59; 60]. The mature miRNA is then loaded into an Argonaute (Ago) family protein inside the RNA-Induced Silencing Complex (RISC) , where it guides RISC for gene regulation [61]. The diverse pathways in miRNA processing have been reviewed in depth in [62] and [48].

MicroRNAs have been found to possess important post-transcriptional regulatory roles in many biological processes and diseases. They bind predominantly to mRNAs that contain partially or fully complementary target sequences [58]. As a consequence, they silence targets through RNA degradation as well as blocking the mRNA translation into proteins. Initially miRNAs were thought to repress protein output with little or no influence on mRNA levels. However, mRNA-array experiments showed that miRNAs decrease the levels of many targeted mRNAs [63]. Perfect pairing of a miRNA with its target site may lead to the destruction of the targeted mRNA through Argonaute-catalysed mRNA cleavage. While this mode of repression dominates in plants, it is extremely rare in mammalian miRNA targets [63].

It is estimated that at least 60% of coding genes are repressed by miRNAs in humans [64]. The `miRBase` database is a searchable database of published miRNA sequences and annotations. The current release (version 21) [65] reports about 1881 precursors and 2588 mature human miRNAs and reports thousands of others across 223 species. However, several studies show that many more miRNAs remain to be discovered, both in well-studied model organisms and in human [66] and therefore the number of published miRNAs is constantly increasing. As an example, the number of miRNA loci annotated in `miRBase` has grown by approximately two-thirds, from 15 172 loci in 142 species (release 16, October 2010) to 24 521 loci in 206 species (release 20, June 2013) [65]. To illustrate the rapid pace of miRNA-related studies, we carried out a keyword search to retrieve publications relevant to miRNA research. We show in Figure 1.8 three indicators of the research status in terms of the number of entries representing hairpin precursor miRNAs stored in `miRBase`, the number of publications reported in

PubMed regarding miRNAs and the number of publications in PubMed with specific reference to miRNA targets.

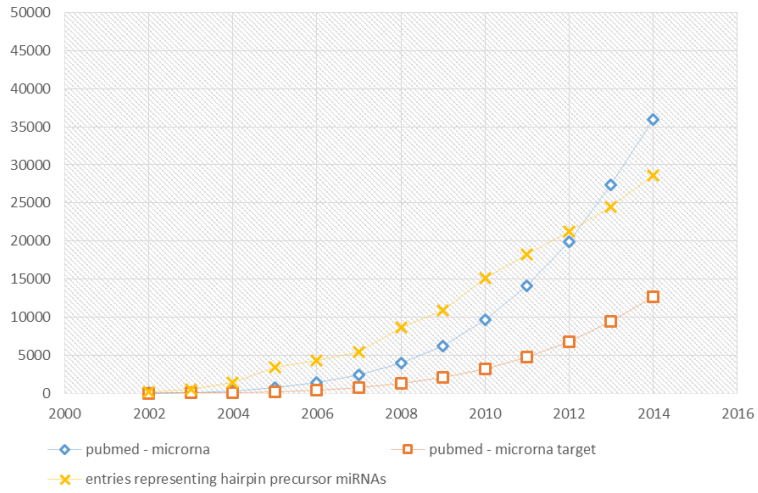


Figure 1.8: MiRNA research timeline showing the annual growth of miRNA-related publications in PubMed and the number of entries in miRBase database.

1.2.2 microRNA and Human Diseases

Although miRNAs have an important role in healthy individuals, increasing work have shown that miRNAs are also implicated in a wide range of human diseases. In fact, miRNAs are now recognized to be involved in different types of cancer, cardiovascular diseases and neurological disorders, where they play a significant role in the disease development, progression and its prognosis [67]. Consequently, miRNAs are becoming a novel class of biomarkers or targets for disease diagnosis and therapy [68]. Experimental approaches have shown that dysregulation of miRNAs is linked to the development of various human cancer [69]. According to [70] and [71], miRNA expression profiles may uniquely identify cancer types while gain or loss of specific miRNAs may function as an oncogene or tumor suppressor. All these evidences commonly suggest that finding miRNA target sites (miRTSs) is expected to be fundamental for novel anti-cancer and viral therapies [72]. Thus, miRNA drugs are currently being tested in clinical trials. However, limits still exist that may prevent miRNA-therapy immediate large-scale exploitation. See

[73] for a detailed review of the personalized medicine field. In the past few years, a number of databases, that provide experimentally supported miRNA-disease associations, have been implemented. A manually curated database entitled `miR2Disease` provides a comprehensive resource of miRNA deregulation in various human diseases [74]. The Human microRNA Disease Database (`HMDD v2.0`) is also a database for experimentally supported human miRNA and disease associations [68].

1.2.3 Experimental Strategies for microRNA Target Identification

MicroRNA targets can be experimentally verified with gene-specific, as well as high-throughput techniques. In gene-specific experiments, by measuring the expression level of the reporter protein or target mRNA, it is possible to draw conclusions whether the miRNA is directly targeting the 3'UTR by translational repression or target degradation respectively, since an inverse relationship between the levels of expression of a miRNA and its target is anticipated.

In vitro reporter gene assays are universally used to confirm a direct regulation of gene expression by a miRNA. To this end, the 3' UTR of the transcript of interest is sub-cloned immediately downstream of the reporter gene luciferase (Photinus or Renilla) or green fluorescent protein (GFP) open reading frame sequence. The construct is then introduced into the cell line expressing the targeting miRNA and the amount of light produced by the luciferase enzyme catalyzing its substrate is measured. A reduction in light output is interpreted as repression of luciferase activity that is induced by the miRNA [75]. Firefly and Renilla luciferases are the most commonly used reporter genes. Alternatively, individual miRNAs can be successively transfected into a cell line that expresses a luciferase reporter containing the 3'UTR of the target mRNA of interest and the expression of reporter gene is measured before and after the introduction of the miRNA to the cell [75]. Such a procedure can provide direct support but fails to identify the specific site of interaction. In fact, to demonstrate a direct miRNA-target interaction in a specific target site, one can measure the gain or loss of miRNA regulation in different constructs with mutated miRTSs [76].

Additionally, gene expression changes can also be experimentally quantified using real time PCR (qRT-PCR). However, miRNAs regulate not only mRNA expression but also protein levels. Therefore, monitoring miRNA effect at the mRNA level may identify targets regulated by target degradation and may leave out targets regulated by translational repression. Thus, the analysis of the effect on target gene expression can also be done at the protein level using western blot to compare protein expression given the presence or absence of the miRNA [76]. Alternatively, the more time-consuming northern blot analysis can be used in the cases where different isoforms of the target are expressed. The disadvantages of these last measures is that they fail to distinguish between direct targets (structural interaction) and targets indirectly down-regulated if they are a part of a miRNA-mediated regulatory module controlled by the direct miRNA targets (indirect targets) [76; 77].

These gene specific experiments are still time consuming and with the increasing number of new miRNAs and their potential targets, high-throughput experimental strategies were developed for large-scale analysis of miRNA targets and their biological function. In fact, since miRNAs act by inhibiting translation and/or promoting degradation of their targets, the approach consists of transfection of specific miRNA mimics or inhibitors into the cells followed by high-throughput analysis of mRNA expression by microarray or high-throughput sequencing or proteomics [78]. As an example, the pSILAC (pulsed Stable Isotope Labelling with Amino acids in Cell culture) method directly measures proteins level changes induced by overexpression of miRNAs using miRNA transfection or endogenous miRNA knockdown [79]. The same technique is used to measure mRNA level changes in microarray experiments [80]. However, none of these high-throughput techniques allows distinction between direct and indirect targets and only gives indirect evidence about specific miRNA-target interactions. HITS-CLIP (HIgh Throughput Sequencing by CrossLinking and ImmunoPrecipitation) is a powerful technique capable of providing an extensive insight to the location of miRNA targeting within an mRNA. However, this technique can only identify a targeted region ($\approx$ 100 nt) as opposed to a specific target site [76]. This technique was improved with the later version called PAR-CLIP (PhotoActivatable-Ribonucleoside-enhanced CrossLinking and ImmunoPrecipitation). Although the

sites of miRNA-mRNA interaction are determined, these approaches can't identify the specific miRNA-mRNA association experimentally. However, this may be estimated using features commonly found in experimental samples such as the seed complementarity feature.

1.2.4 Problem Formulation

The problem of discovering the targeting genes that miRNAs regulate is of a great importance to developing our understanding of many diseases and viruses. The number of discovered miRNAs is increasing and each miRNA is thought to regulate a few hundred target mRNAs [79; 81]. Moreover, experimental based predictions are time consuming and expensive while most of the real targets are yet to be discovered. Consequently, efficient and reliable computational methods for target prediction are required to generate clues and hypotheses for experimental studies.

In plants, an almost perfect base pair complementarity of the whole miRNA to the mRNA is required, making target identification a simple task [82]. In animals, on the other hand, miRNAs are only partially complementary to their mRNA target sequences [83]. Therefore, computationally identifying miRNA target genes is a challenging problem in bioinformatics. Alkan *et al.* studied the general RNA-RNA interaction prediction problem under different models and proved that it is an NP hard problem [84]. Hence predicted structures need to be simplified by removing pseudo-knots such that secondary structures have no crossed base pairs.

1.3 Important Features in microRNA Target Prediction

Common characteristics found in data with strong experimental support are being identified. In fact, empirical evidences are examined carefully to extract the principles of miRNA target recognition. Computational methods perform the predictions based on these features. In this section, we discuss a few important principles applied in the prediction algorithms.

1.3.1 Seed Match Requirement

Some methods require the target sequence to have a perfect complementarity to a region in the mature miRNA sequence named seed. The seed of a miRNA is defined as the sequence between the second base and the 8^{th} nucleotide of the 5' end. This requirement is justified by the fact that the seed region is believed to be the most evolutionary conserved region of miRNA [85] and that most of the experimentally validated targets have been observed to have a high degree of complementarity to the seed sequence of the miRNA. The importance of the seed match is further emphasized by a study of the crystal structure of the silencing complex RISC that shows that the seed region is tightly bound to the complex [86] and also by the observation that miRNAs with almost identical sequences at their 5' ends, which form miRNA seed families, share targets [87]. The existing methods which use the seed match, require as few as 6-nucleotide matching at the seed region [88]. There are possible 6-mers (positions 1–6, 2–7, and 3–8 from the 5' end of the miRNA), 7-mers (positions 2–8 and 1–7 from the 5' end of the miRNA), and 8-mer (position 1–8 from the 5' end of the miRNA) matches in the seed. Otherwise, a 6-mer match to position 3–8 is called an "offset 6-mer seed" because of its position and a marginal effect on repression [88]. Prediction methods with seed-pairing criteria have their false positive rate and their time complexity reduced [89]. Nevertheless, some experimentally validated miRNA targets that do not have a perfect seed match have been identified. Pairing to the 3' region of the miRNA can sometimes supplement seed matches to enhance target recognition, or it can compensate for mismatches in the seed. Such sites are known as '3'-supplementary sites' and '3'-compensatory sites' respectively [88]. Moreover, a class of 'centered sites' described as 11-12 contiguous base pairs to the center of the miRNA can compensate for the weaker seed complementarity leading to the assumption that non-seed nucleotides may have a significant free energy contribution [90]. Therefore in order to avoid missing true targets, some tools give matches at some positions a different weight, leading to a weighted 5' end scheme.

1.3.2 Sequence Conservation

The miRNA and its targets are shown to be widely conserved across a wide range of species [64]. Therefore, some computational approaches applied to predict miRNA targets rely on sequence search similarity between different species, considering binding site evolutionary conservation. The conservation criteria can be incorporated into predictions in a variety of ways. While some tools require target sites to appear at the same position in a cross-species UTR alignment [89], others require the binding to only occur at overlapping positions in the alignment of the 3'UTR sequences [91]. Moreover, the number of species involved in UTR alignment differs with some computational tools requiring only conservation between human and rodent while others allow extensive conservation involving more than two species [92]. Considering binding site evolutionary conservation allows high prediction accuracy [93]. However, further studies have shown that not all target sites are necessarily conserved. For example, in [92], the analysis of 84 experimentally verified interactions provided by the `DIANA-TarBase` database gives 23 interactions with no conserved target sites. Thus, using the conservation criteria may come at the expense of missing real targets. Additionally, binding site conservation evaluation involves multiple sequence alignments over the different species and this can be computationally heavy.

1.3.3 Thermodynamics

Similar to the secondary structure prediction problem, early tools used a base pair counting model to predict interactions between mRNA and miRNA. The drawback of using this model is that base pairs maximizations do not necessarily lead to the most stable structure as it may create a structure with many interior loops or hairpins, which are energetically unfavorable. Therefore, the main focus of our study will be on methods that are based on thermodynamic stability of the duplex structure, considering free energy minimization. In fact, the most popular miRNA target prediction tools rely on traditional RNA secondary structure prediction, where the change in Gibbs free energy ΔG_{duplex} of the miRNA-mRNA duplex structure is calculated and used as a determinant component for evaluating their interaction. The lower the free energy of the two paired RNAs, the

stronger is the binding and the more energy is needed to disrupt this duplex formation. Consequently, a miRNA has a strong affinity to a target mRNA when the corresponding duplex has a low ΔG_{duplex}.

1.3.4 Target Site Accessibility

Some tools avoid intra-molecular base pairing by omitting the computation of folded structures within the monomers and rely almost exclusively on the free energy change of the duplex formation ΔG_{duplex}. This assumption that the mRNA is in linear form certainly reduces the computational complexity. However, recent studies suggest that this assumption describes only part of the binding process and that prediction tools can be improved by incorporating the folded structure of the mRNA into the prediction algorithm [24; 94]. In reality, either the binding site must not be involved in any base pairing with other parts of the same mRNA or there should be an energetic penalty ΔG_{open} for freeing base pairing interactions within the mRNA in order to make the target accessible for the binding. This energetic cost has to be considered in the total hybridization energy [24; 95]. Methods considering the accessibility of the binding site instead of its conservation provided an alternative way of increasing precision [24]. This feature is used in a variety of ways. The standard assumption in miRNA target predictions is that the functional state of the mRNA is the MFE structure. However, recent literature argues in favor of the existence of multiple active RNA conformations instead of a unique MFE conformation as the single biologically active state [96]. Long *et al.* [97] overcame the limitation of using a single MFE structure in predictions, by averaging over 1,000 structures sampled from a statistically representative sample from the Boltzmann-weighted ensemble of RNA secondary structures by using the stochastic sampling method `Sfold` [98]. In [24], the authors compute the accessibility in relation to the probability that the target region is unpaired in thermodynamic equilibrium and, additionally, based on the ensemble of all possible structures in thermodynamic equilibrium, where `RNAfold` [22] is utilized. Marín and J. Vaníček [99] argue that considering only the MFE structure neglects the possibility that the miRNA binds to a 3'UTR structure with a slightly higher energy than the MFE structure, but with

better accessibility. The authors compute pair probabilities from the canonical ensemble of secondary structures generated by `RNAplfold` [100] in order to find accessible stretches of nucleotides in the seed matches. They achieve comparable or better results obtained with MFE structures only. A few other computational approaches compute the local accessibilities i.e. the probabilities that sequence intervals are single-stranded in thermodynamic equilibrium. `RNAup` tool [101] uses the local folding algorithm `RNAplfold` [102] to compute the accessibilities of all intervals of an RNA molecule. `RNAplfold` algorithm has later been rewritten to use a more efficient version, which has the same complexity as `Sfold` (requires cubic time) [103].

1.3.5 Expression Levels of RNAs

Within the past few years, analyzing concentration levels of miRNAs and their putative targets has become a major topic in miRNA research. In [104], the authors provide experimental evidence that the typical number of gene copies present in a single cell lies between 5–20, with most genes having less than a 100 copies. Individual miRNAs are likewise considered to vary widely, with a few tissue-specific species present more than 10,000 copies per cell [105]. Consequently, miRNAs are more abundant than mRNAs with an average of 500 copies per cell. Therefore, the miRNA abundance can help to explain the co-regulation of a target mRNA by several miRNAs, and the regulation of multiple mRNAs by a single miRNA. Ragan *et al.* [106] published results on miRNA target predictions that utilize information about miRNA and mRNA concentration levels. For the miRNA target prediction tool `TargetScan`, Garcia *et al.* [107] demonstrate how predictions may improve if target abundance is accounted for in binding scores. Other implementations such as `Sylamer` have explored the combined analysis of miRNA and mRNA expression data to increase the accuracy of miRNA target prediction and to uncover miRNAs that actively regulate mRNA expression following a miRNA experiment [108].

1.3.6 Location of Binding Region

Although functional miRNA sites are thought to be predominantly located in the 3'UTR of mRNAs, studies where miRNAs have been reported to regulate gene expression through their binding to the 5'UTR region of target genes have emerged [109; 110]. Seed sites located in the coding sequence of mRNAs can also cause downregulation [111]. However, the implication of these interactions remains obscure due to their smaller impact on mRNA stability compared to miRNA-target interactions that involve 3'UTRs [112].

1.4 Overview of Computational Methods for microRNA Target Prediction

The first miRNA target prediction method was published in 2003 by Stark *et al.* [113]. This early algorithm for miRNA targets was based on the few known miRNA-target pairs, comparison of *Drosophila melanogaster* and *Drosophila pseudoobscura*, and minimal assumptions regarding miRNA-target interaction. The algorithm successfully predicted new targets that the authors validated experimentally. Since then, several other computational methods have been published and made available to users. The prediction algorithms are mainly divided into two main groups: data-driven and rule-based approaches [114]. Data driven tools collect different types of features and use machine learning techniques to find the feature patterns shared by true miRNA-target interactions. Such tools are affected by shortage of reliable training data and the lack of published works related to thoroughly experimentally refuted targets limits the validity of their results. On the other hand, rule based algorithms consist of a set of rules tested according to a particular order, considered as filtered steps [114]. In this section, some of the most commonly used algorithms for human miRNA target prediction are reviewed. Our review will be mainly focused on rule based algorithms, although some of the machine learning approaches will be discussed.

1.4.1 miRanda/miRanda-mirSVR/microRNA.org

`miRanda` is one of the earliest developed target prediction tools implemented in 2003 [115]. The algorithm consists of a three-step analysis. It first calculates a score using a position weighted dynamic programming algorithm to assess the complementarity between miRNA and its target. The scoring method favors complementarity between the 5'end of the miRNA i.e. the seed region and the 3'end of the mRNA. However this tool does not require perfect seed complementarity. As a second step, `miRanda` estimates the free energy of the duplex by using `RNAfold` [11] from the Vienna package. To comply with the input of `RNAfold`, the miRNA is concatenated with the potential binding site, inserting an artificial 8-bit linker sequence between them formed by the character 'X'. Finally, conservation in related genomes is used in order to validate predicted target sites and to reduce false positives. However, this tool also allows looking for non-conserved target sites. As a final outcome, this tool classifies a site as a real binding site if its free energy is less than a cut-off value. Although the algorithm was originally used to identify targets in *Drosophila*, subsequent versions were applied to predict miRNA-target in humans [116]. `miRanda` provides a searchable precompiled dataset and the latest release from 2010 includes the human, mouse, rat, fruit fly and Nematode species. It also provides a freely downloadable version of their sourcecode in order for the user to run it locally with their own pairs of miRNA and target sequences. Later on, Betel *et al.* [117] designed a new algorithm called `mirSVR` for scoring and ranking the efficiency of `miRanda`-predicted miRTSs by using supervised learning on mRNA expression changes following miRNA transfections. This machine learning method combines target site information and contextual features into a single integrated model, and uses support vector regression (SVR) to train on a wide range of features, including secondary structure accessibility of the site and conservation. Similar to `miRanda`, `miRanda-mirSVR` provides a searchable precompiled dataset. However, the program's results and data resources are not always updated and therefore the latest version of `miRBase` is not used for the predictions.

1.4.2 TargetScan

`TargetScan` is one of the earliest developed target prediction tools, implemented in 2003 [85]. The algorithm primarily takes the perfect complementarity rule of the seed region to the putative target sequence into account as a first filtering step. It also mainly uses the conservation criteria for filtering. Predicted binding sites are then evaluated thermodynamically using `RNAfold` [11] from Vienna RNA folding package and a final score is computed. `TargetScanS` [89] is an improved and simplified version of this tool that does not consider free energy and limits the miRNA complementarity to a six-nucleotide seed match and a match at position 1 of 3'end. Conservation is finally required between five species in order for the binding site to be valid. However, non-conserved sites can also be predicted as an option. Later versions relaxed the seed match requirement and identify sites with mismatches in the seed region that are compensated by conserved 3' pairing [64] and centered sites [90]. Furthermore, a number of additional features have also been integrated. In particular, a multiple linear regression, trained on 74 filtered datasets, was used to integrate determinants such as seed-pairing stability (SPS) and target-site abundance (TA) [107]. `TargetScan` web interface allows the user to search by miRNA name, gene name, or from broadly conserved, conserved, or poorly conserved miRNA families across several species. The latest release 7 dates from 2015 [118] and includes searchable lists for human (TargetScanHuman), mouse (TargetScanMouse), worms (TargetScanWorm) , fruit fly (TargetScanFly) and fish (TargetScanFish) species. TargetScanHuman considers matches to human 3' UTRs and their orthologs, as defined by `UCSC` whole-genome alignments [119]. `TargetScan` Perl scripts are also available to download, the last version dates from 2015.

1.4.3 MicroCosm

`MicroCosm` Targets (previously known as `miRBase` Targets) is a web resource containing computationally predicted targets for miRNAs across many species [115; 120]. `MicroCosm` uses the `miRanda` algorithm to identify potential binding sites. The current version 5 uses dynamic programming alignment to identify highly complementary sites, which are scored between 0 and 100, where 0 repre-

sents no complementarity and 100 complete complementary. Strict complementarity to the 5' seed region is required by the algorithm. Furthermore, target sites selected by `miRanda` are passed through the Vienna RNA folding routines [11] to evaluate their thermodynamic property. Finally every potential target site is checked to see whether the site is conserved in orthologous transcripts from other species. The last release of the database (v5) dates from 2008.

1.4.4 RNAHybrid

`RNAHybrid` [121] directly predicts optimal and suboptimal secondary structures for the hybridization of miRNA and a large target RNA. The algorithm finds the energetically most favorable hybridizations of a small RNA to a large RNA. The calculation is simplified by allowing only inter-molecular base pairs. Using dynamic programming technique, bulge loops and internal loops are restricted to a maximum length. `RNAhybrid` uses perfect seed match by default (although this can be user defined) to find all possible binding structures and picks the structure which gives the MFE. This MFE along with its p-value are compared to user-defined cut-offs in order to classify the given 3'UTR as a target to the input miRNA. The energy model used in `RNAhybrid` is essentially equivalent to the classic RNA folding algorithm of Zuker [12] where only interior loops are allowed. In case of long sequences, it is possible that many hybridization sites reach significant interaction energy and therefore `RNAhybrid` reports suboptimal hits. A later version of the tool have been implemented, where a few features, such as the possibility to disallow G:U base pairs in the seed region, and a seed-match speed-up, which accelerates the program by a factor of 8, have been included [122]. RNAhybrid was originally available through a web command line downloadable tool. However, the program can now also be used as a web service for remote calls, thus eliminating the need for a local installation.

1.4.5 PicTar/doRiNA

The Probabilistic Identification of Combinations of Target sites (`PicTar`) method, first implemented in 2005, allows the identification of targets for both single miRNAs and combinations of miRNAs. Input to `PicTar` consists of a fixed search

set of co-expressed miRNAs and multiple alignments of RNA sequences (typically 3' UTRs). Outputs are scores that rank genes by their likelihood of being a common target of members (subsets) of the search set and probabilities for the predicted binding sites in each UTR [91]. As a first step, the algorithm looks at perfect seed matches, perfectly Watson-Crick base paired stretches of ≈ 7 nt, to the given miRNA among a group of 3'UTRs. In the case of imperfect matches, the free energy is computed and compared to a threshold binding energy. Once the probabilities for each subsequence of the RNA sequence to be a binding site for the miRNA are fixed, a score is computed based on the previous steps, taking into account the conservation across species. This tool is represented as a searchable precompiled dataset available via the novel database `DoRiNA` [123], where the most recent `PicTar` predictions for all species date from 2014 [124].

1.4.6 PITA

Probability of Interaction by Target Accessibility (`PITA`) [24] uses target-site accessibility as the major feature for miRNA target prediction. The algorithm first scans the mRNA sequence for potential target sites by searching near perfect seed matches. Then for each binding site within the mRNA, a thermodynamic model is used to compute the accessibility energy score $\Delta\Delta G = \Delta G_{duplex} - \Delta G_{open}$, that is the difference between the energy of the duplex ΔG_{duplex} and the energy required for making the target region accessible for miRNA binding ΔG_{open}. The tool `RNAduplex` [125] from the Vienna package is used to predict the MFE structure of the duplex ΔG_{duplex}. Computing the accessibility is directly related to the probability that the target region is unpaired in thermodynamic equilibrium and is based on the ensemble of all possible structures in thermodynamic equilibrium rather than a single MFE structure, using the `RNAfold` tool[22]. The disruption energy ΔG_{open} is determined by computing the free energy difference between the native mRNA secondary structures and the same structures with the target region, including additional nucleotides upstream and downstream, required to be unpaired. As a final step and in the case of multiple sites for a single miRNA, the energy scores are appropriately summed to form a score for the total interaction energy of the duplex. `PITA` is available as an online user interface, where the user

can choose between the fly, worm, mouse and human organisms and can enter any seed match parameters. Additionally, PITA webpage provides precompiled dataset and the latest update of PITA Catalog version 6 dates from 2008. Users have also the option to download PITA executable in order to run the tool on the command line.

1.4.7 STarMir

STarMir[97] RNA target prediction tool is based on a two-step nucleation expansion model. In the two-step model, hybridization nucleates at an accessible target site, and then the binding elongates to form the complete miRNA-target duplex. To overcome the limitations of using a single MFE structure in predictions, Sfold algorithm [98] is used to get a sample of 1,000 secondary structures from the entire Boltzmann ensemble to compute target accessibility profiles for the putative target sites.

- **Nucleation stage**
  This stage is related to the assumption that nucleation requires a gain in free energy from base-pairing at the nucleation site that is greater than the energy cost for the translational and rotational entropy loss when two RNA strands are fixed in a conformation by intermolecular base-pairing. The energy cost for the translational and rotational entropy loss is called initiation energy $\Delta G_{initiation}$. The standard setting in STarMir is $\Delta G_{initiation}$ = 4.09 kcal/mol. The first step in this stage is to find potential target sites. With the help of a window of length four nucleotides, the accessibility profile shows the probability that the four consecutive nucleotides starting at the indicated nucleotide are all single stranded. A potential site is a block that has a probability of 0.5 or greater of being unpaired. As a second step, for each potential site, the best binding with the miRNA is computed. For each secondary structure $i, 1 \leq i \leq 1,000$, compute the nucleation potential $\Delta G_{N,i}$ over all structures from the sample structures and then the average is calculated. In fact, the algorithm checks for each structure in the sample how many four consecutive nucleotides are free within the binding site and then calculates the binding of these four nucleotides and takes the smallest

of all, i.e. the strongest binding. This energy value will represent the structure and finally, an average of these values is computed and given as a nucleation potential ΔG_N. The requirement for further processing of the particular miRNA-mRNA interaction is given by:
$\Delta G_N + \Delta G_{initiation} < 0 kcal/mol$.

- **Elongation stage**
  Once a target site has passed the nucleation threshold, the total interaction energy between the miRNA and the mRNA is computed, using a similar method to the model used in `PITA`. The tool `RNAhybrid` [121] is used to compute the MFE structure of duplex ΔG_{hybrid} which is required to be less than a threshold of -14 kcal/mol. The difference between the free energy of the duplex ΔG_{hybrid} and the free energy needed to disrupt the structure of the target site ΔG_{open} is computed and gives the main score for this potential binding site similar to the $\Delta\Delta G$ score for `PITA`. ΔG_{open} is computed as a difference between the native mRNA structure free energy and the free energy of the same structure with the target site unpaired. The average over all the structures is computed.

In the case of multiple sites, the scores are linearly summed and compared to a threshold (-10 kcal/mol). `STarMir` adds the extension to the coding region and 5'UTR to predict the secondary structure as an option. Moreover, there is no requirement for a seed match in contrast to most other prediction methods. Therefore, this tool is generic and not specific to the miRNA target prediction. A later version, implemented in 2013, introduced logistic prediction models developed with miRNA binding data from CrossLinking and ImmunoPrecipitation (CLIP) experiments [126]. The model is also utilizing evolutionary conservation information for predictions if the user specifies the sequence using the RefSeq ID. `STarMirDB` [127], a web searchable database currently including all predicted binding sites by the models for miRNAs and mRNAs in the HITS-CLIP and V-CLIP studies, with an indicator showing whether a site is supported by the CLIP study, has also been developed, for human mouse and worm species. `STarMir` web server [128] allows users to submit their own miRNA and mRNA sequences for prediction of binding sites by the models.

1.4.8 DIANA-microT/DIANA-microT-CDS

`DIANA-microT` [129] is one of the most widely used tools since its initial launch in 2009. The initial algorithm is based on the dynamic programming to predict potential binding sites and uses a 38-nucleotide window that progressively moves across the 3'UTR sequence one nucleotide at a time. The putative pair are first assessed whether there were at least three consecutive canonical pairings between the two sequences and then the MFE of the potential site is calculated. Finally, the algorithm filters weaker bindings based on specific rules focusing on the number of consecutive base pairs towards the 5' end region of the miRNA as well as the 3'end and the size of allowable bulges in the miRNA or its cognate mRNA. In [130], a later version was implemented with a change in the binding rules, focusing more on the alignment between the seed region of the miRNA and the target sequence. Target sites are then scored according to their binding category and degree of conservation in other species and an overall miRNA-target gene (miTG) score is calculated through the weighted sum of all target sites' scores lying on the 3'UTR. The algorithm also allows looking for non-conserved target sites. Later on, a new machine-learning version of the microT algorithm, `DIANA-microT-CDS`, has been specifically designed to identify miRNA targets both in 3'UTRs and in coding sequences (CDS) [131]. `DIANA-microT` web server v5.0 hosts the most recent version of `DIANA-microT-CDS` algorithm [132] implemented in 2013, and is using data from Ensembl version 69 and `miRBase` version 18, and currently hosts miRNA target predictions for *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster* and *C. elegans*.

1.4.9 Sylamer/SylArray

`Sylamer` is a prediction method for detecting microRNA target from expression data [108]. The algorithm quantifies the over or under representation of miRNA seed matches in a sorted list of genes from an expression experiment. The significance of enriched binding sites is calculated using hypergeometric p-values. `SylArray` [133] provides a database and a user interface to the `Sylamer` algorithm. It allows the user to upload an ordered gene list obtained from a microarray or other high-throughput miRNA experiment. `SylArray` utilizes cu-

rated sets of 3'UTRs to attach sequences to these genes and then applies the `Sylamer` algorithm for detection of miRNA signatures in those sequences. The graphical output is a landscape plot of the enrichment p-values for miRNA seed complementary sites in 3'UTRs, calculated in incremental parts of the submitted ranked gene list [133].

1.5 Online Resources for microRNA Research

In this section we present a brief overview of the most widely used databases and repositories, which offer fundamental data resources to any miRNA research process. Some prediction tools provide precompiled predictions on their website such as `PITA`, `TargetScan` and `Pictar`. Others provide web interfaces such as `STarMir` or allow users to download a version of their program such as `miRanda`. To accurately measure the performance of a prediction tool, it is important to have a sufficient number of experimentally validated and refuted miRNA-target interactions. An important source of experimentally validated miRNA-target interactions is the `DIANA-TarBase` database provided by the Diana lab. The latest version v7.0 from 2014 includes more than half a million miRNA:gene interactions, curated from published experiments implemented utilizing 356 different cell types from 24 species [134]. For each interaction, it provides a detailed description of the gene and the miRNA, its related publications and the experiments used for its validation. Another resource for experimentally validated miRNA targets is the `miRecords` database. The latest version, dated from 2013, hosts 2705 records of interactions between 644 miRNAs and 1901 target genes in 9 animal species. Among these records, 2028 were curated from low-throughput experiments [135]. The `miRTarBase` provides the most updated collection by comparing with other similar, previously developed databases [136]. The `starBase` database [137] collects data provided by high-throughput CLIP-seq (PAR-CLIP, HITS-CLIP, iCLIP, CLASH). This type of data consists of sites of interaction for the mRNA-miRNA-Argonaute complex on a transcriptome-wide scale. The last version v2.0 [138] identifies RNA-RNA and protein-RNA interaction data from 108 CLIP-Seq generated by 37 independent studies.

Furthermore, several human genome annotation databases have been devel-

oped. These include RefGene (RefSeq Gene [139]), Ensembl [140], and the UCSC annotation database [141]. For most cases the data is the same between these different repositories. However, in [142], the authors demonstrate that the choice of a gene model may have a dramatic effect on both gene quantification and differential analysis. miRBase database [65] is the main used repository for published miRNA annotation and nomenclature. microRNA.org [143] provides a collection of miRNA expression profiles in various tissues and cell lines while the Expression Atlas [144] provides information on gene expression patterns under different biological conditions.

1.6 Comparison of Prediction Methods

Target prediction tools certainly facilitate target identification as they are used to pre-select the putative targets of a given miRNA. However, biologists are facing the problem of how to select the best tool to use based on the needs of a particular experiment. Moreover, the vast number of predicted targets makes it a challenge for scientists to choose which interactions are worthy to validate experimentally, and which ones will have a major impact in a given biological pathway under study. Some miRNA target computational methods are available as web servers allowing users to search precomputed prediction results. These tools are convenient and usually easy to use. However, the annotation databases are updated frequently and especially for miRNAs, the differences between versions are significant and therefore the available predictions should be updated regularly. Furthermore, some web servers do allow predicting for novel mRNA and miRNA sequences, allowing users to provide mRNA and miRNA sequences. These include STarMir and PITA web servers. Other computational methods are available as stand-alone packages, allowing the user to run their own predictions on their local machine. This is particularly useful when predicting for a large number of mRNAs and miRNAs. When predicting the target site, the methods that can only predict target genes cannot be used. The performance of computational tools can be measured by standard measures like sensitivity also called true positive rate (TPR), specificity and false positive rate (FPR).

$$Sensitivity = TPR = \frac{TP}{TP+FN}$$

TP (true positives) is the number of predicted miRNA-target interactions that really exist. FN (false negatives) is the number of miRNA-target interactions that do exist but are not predicted. Thus, sensitivity is the percentage of correctly predicted targets out of total correct interactions.

$$Specificity = \frac{TN}{TN+FP}$$

$$FPR = \frac{FP}{FP+TN} = 1 - Specificity$$

FP (false positives) are the experimentally rejected miRNA-mRNA interactions that are wrongly predicted by the tool. TN (true negatives) are also important, that is is the number of non-existing miRNA-target interactions that are correctly not included in the predictions. Subsequently, specificity is the relation of the number of correctly not predicted interactions that do not exist to the number of all experimentally refuted interactions. Both sensitivity and specificity have to be maximized to achieve a good performance and the FPR should be minimized. A suitable method to evaluate the relation between the sensitivity and the specificity is a Receiver Operating Characteristic (ROC) analysis. The ROC curve is a plot of the sensitivity versus the FPR, which helps to find the optimal balance between these two measures. The area under the curve (AUC) is a comparable value for the overall performance and it ranges between 0 for a method that does not predict TP and 1 for a perfect predictor, with 0.5 denoting a meaningless result. Several computational methods were benchmarked using datasets identified by high-throughput experiments. In [117], Betel *et al.* compared `mirSVR` against a number of existing target prediction algorithms using a large panel of independent miRNA transfection and inhibition experiments as test data. Their results show that `mirSVR` strongly outperforms the alignment-based (`miRanda`) and energy-based (`PITA`) scores for the task of ranking single-site genes by their downregulation (upregulation) in response to microRNA transfection (inhibition). However, these methods were not trained on genome-wide expression data and therefore were not expected to perform as

well as supervised approaches such as mirSVR. On the other hand, mirSVR performs better than TargetScan's context score in 21 out of the 25 test sets, which constitutes a statistically significant improvement. In [127], the performance of STarMir's CLIP-based logistic model was tested by both intra-dataset validation and inter-dataset validation. Liu *et al.* compared STarMir predictions with predictions available from TargetScan, mirSVR, PITA and RNA22 [145]. Logistic models trained on five different CLIP datasets all have substantially higher TPR than TargetScan and mirSVR, RNA22 with comparable FPR, and lower FPR than PITA, RNA22 with comparable TPR. Additionally, for inter-dataset validation, the authors noticed an improvement in predictions. In [118], Agarwale *et al.* compared predictions of TargetScan7 to several miRNA target prediction tools, including DIANA-microT-CDS, miRanda-MicroCosm v5, mirSVR, PicTar2 and PITA Catalog v6, using the results of seven microarray datasets. The authors tested how well each of the methods predicted the repression of mRNAs with at least one canonical 7-8 nt 3' UTR site and found that TargetScan's context++ model, DIANA-microT-CDS and miRanda-miRSVR were among the most predictive tools. In [131], the prediction sensitivity and precision for several methods is tested on a dataset identified by the pSILAC method [79]. The DIANA-microT-CDS program exhibits the highest sensitivity at any level of specificity in comparison with six other programs including miRanda, Pictar and TargetScan 5.0.

Additionally, several surveys were proposed to analyse and compare computational methods using independent datasets. In [146], to test the sensitivity of target prediction algorithms, the authors used HITS-CLIP datasets for human, mouse and *C. elegans* downloaded from the starBase database. miRanda algorithm (applied on 3' UTR sequences) shows the maximum coverage of miRTSs (66%) followed by PITA, TargetScan and Pictar. In [147], the authors give a recent comprehensive overview and assessment of computational prediction of miRNA targets in animals. The authors compared the performance of seven predictors on four benchmark datasets and found that although certain methods, like TargetScan and miRmap [148], offer high overall predictive quality, there is no a universally best predictor. For instance, PicTar and MirTarget2 [149] provide predictions with high specificity and low number of FP.

It is important to highlight that none of the existing tools could capture all

true targets and the FPR is still relatively high. Moreover, the performance of a method is not consistent across different datasets. Thus, how to obtain reliable and comprehensive results is a long-standing challenge for miRNA target prediction. As a result, several tools have been integrated into one body of information to form either a database such as `miRecords`' Predicted Targets component [135], an integration of predicted miRNA targets produced by 11 established miRNA target prediction programs, or into a single miRNA target prediction algorithm such as the `SMILE` tool [150], which integrates the outcomes of individual prediction tools with the aim of surpassing the performance of the individual tools. A detailed review of some of the existing integrated data resources can be found in [151].

1.7 miRNA-related Polymorphisms as Biomarkers

Polymorphisms are DNA sequence variations that may exist in several forms. The most common forms are Single Nucleotide Polymorphisms (SNPs) which involve a change in one single nucleotide. SNPs typically have two alleles at the specific base position in the genome i.e. within a population there are two commonly occuring nucleotide variations [152]. For a given SNP, the allele that was either first discovered or is the most common is considered the reference against which all other forms are compared. This reference form is called the wild-type allele [153]. The allele that is observed to have the lowest frequency at a locus in a particular population, is called the minor allele.

miRNA-related polymorphisms include SNPs located within miRNA sequence or miRNA-binding target sites. These are defined as miRSNPs and are typically located at the seed sequence and the 3'UTR of mRNAs, respectively. Small insertions and deletions (INDELS) in miRNA sequences and miRTSs are also considered as the second largest class of genetic variants after SNPs [154]. miRNA-related polymorphisms are increasingly thought to play a significant role in pathological dysregulation of gene expression. In fact, these polymorphisms have been associated with many human diseases, including cancer [155], diabetes

[156], inflammatory diseases [157], Parkinsons disease [158] and many other diseases. Consequently, identifying functional miRNA-related SNPs are of interest for disease-related studies.

A genome-wide association study (GWAS) is an approach that involves scanning markers across complete sets of genomes of many individuals to find genetic variations associated with a particular disease. GWASs have successfully discovered hundreds of novel genomic loci that influence human disease susceptibility and therefore helped identify genetic risk factors for common diseases in the population [152]. Alleles that are associated with higher risk of a particular human disease are called risk alleles. The current build 144 of `dbSNP` database [159] include around 100 millions validated human SNPs. GWASs are based upon the principle of linkage disequilibrium (LD) at the population level. LD describes the degree to which alleles at two or more nearby loci are correlated to each other within a population [152]. The LD structure was investigated in the HapMap project [4] and the outcome was a list of SNPs that captured most of the common genomic variation in a number of human populations. A brief introduction to the the early history of GWASs is given in [160], where the authors list some of the first discoveries made through this experimental design.

A recent review of miRSNPs as biomarkers in cancer management and research is given in [155]. miRNA dysregulation in cancer was first reported in 2002 when miR-15 and miR-16 were shown to be involved in Chronic Lymphocytic Leukemia (CLL) pathogenesis [161]. A subsequent study demonstrated that a variation in the primary sequence of the miRNA cluster encoding miR-16-1 and miR-15a results in reduced mature expression of miR-16-1 and miR-15 *in vitro* and *in vivo*, and is associated with deletion of the chromosome region containing this miRNA cluster [162]. This DNA variant was found in 11 of 75 patients with CLL but not observed in 160 subjects without cancer [162].

miRNA-related SNPs were first reported to affect phenotype in 2005 when a mutation in a miR-189 binding site of SLITRK1 was found to be associated with Tourette's syndrome [163]. The authors showed that the 3'UTR SNP destroys a target site for miR-189, making this work the first evidence that a 3'UTR SNP could inhibit miRNA binding [163]. The first work to report that a 3'UTR SNP could create an illegitimate miRNA target site was published in 2006 [164].

The authors showed that a SNP in the 3'UTR of myostatin caused muscular hypertrophy in Texel sheep [164]. The SNP creates an illegitimate target site for miR-1 and miR-206, which are highly expressed in skeletal muscle and specifically target the variant allele.

Since these initial observations, a number of studies have used systematic sequencing and *in silico* approaches to identify SNPs in miRNA target genes. As an example, it has been shown that a SNP in the 3'UTR of KRAS gene, located in the binding site of miRNA let-7, weakens its inhibition and increases the risk of non small cell lung cancer [165].

Polymorphisms within miRNA genes have been reported to be relatively rare, with only approximately 10% of human pre-miRNAs having documented SNPs, and less than 1% of miRNAs having SNPs in the functional seed region [166]. As a consequence, the information about miRNA seed region polymorphisms has received much less attention compared to the SNPs that occur in the miRNA target site [167]. As an example, a study discovered that a mutation in the seed region of human miR-96 was responsible for nonsyndromic progressive hearing loss [168]. A detailed review listing some of these disease-associated mutations and their effect on miRNA regulation is given in [169].

Several online databases have been developed storing SNPs in miRTSs and/or miRNA sequences. `mirdSNP` database [170] is a database including disease-associated SNPs (dSNPs) on the 3'UTRs of human genes manually curated from `PubMed`. `PolymiRTS` (Polymorphism in microRNAs and their Target Sites) is a database of naturally occurring DNA variations in miRNA seed regions and miRTSs [171]. There also exist two other databases with a similar purpose to `PolymiRTS`: `MicroSNiPer` [172] and `Patrocles` [173], which integrate SNPs, phenotype, and expression data. However, the number of discovered SNPs is still increasing following the same pace as the rapidly growing genomic data and therefore these databases should be updated on a regular basis. Sethupathy and Collins [174] critically assess several genetic association studies related to miRNA bindings and the potential impact of SNPs in bindings regions. The authors highlight the importance of follow-up functional experiments for a deeper understanding of the real effect miRNA target site variations may have on the development of various human diseases.

1.8 Organisation of the Thesis

The present work is an investigation into the impact of disease-associated SNPs on the thermodynamics of the mRNA-miRNA interaction and accessibility of the target and the role of computational methods in predicting this impact. We are particularly interested in exploring how these SNP-effect predictions could be more accurate if accessibility features related to RNA secondary structures different from MFE conformations are taken into account. Chapter 2 introduces our manually curated dataset and the main motivation of our work. The rest of the thesis is organized as follows. The problem of miRNA bindings to metastable secondary structures in the context of SNPs and mRNA expression levels will be studied in chapter 3. We showed that the number of metastable structures and the features of miRNA bindings to metastable conformations can provide additional information supporting the differences in expression levels of mRNAs and their corresponding SNP-infected variants. In chapter 4, a novel target prediction tool that incorporates metastable secondary structures with low energy levels into predictions is introduced. The algorithm of the proposed method is discussed and a summary of the test results is given. In the last chapter, we summarize the thesis and its contributions and finally discuss the limitations of our approach and future directions.

Chapter 2

Disease-associated SNPs Within miRNA Target Sites Dataset

The aim of this chapter is to introduce the main motivation of our work on how to improve detecting the effects of the miRSNPs on miRNA bindings. We first introduce the dataset that will be used throughout our work. This dataset includes 34, manually curated, disease associated SNPs within miRNA target site cases. We analyzed the relevant publications and we extracted the common framework followed in most of these studies and listed the most frequently used miRNA target prediction tools. We then tested the performance of the available databases on our dataset of 34 disease associated SNPs. Three categories of miRNA-target gene related research databases were of particular interest: the databases which offer repositories for experimentally validated miRNA-target interactions, the databases which provide direct access to miRNA-target predictions from several well established tools and finally databases offering SNP-effect predictions. A comparison between these databases is given for each category. We finally compared results of two major computational approaches to miRNA target ranking prediction: conservation feature using `TargetScan` tool and target site accessibility feature using `PITA` and `STarMir` tools.

2.1 Background

miRNAs are short non-coding RNAs known to possess important post- transcriptional regulatory roles. They bind to mRNAs that contain specific complementary target sub-sequences, and this way blocking the mRNA translation into proteins [58]. Hundreds of targeted genes associated with cancer, cardiovascular disease, viral infections and other diseases have been experimentally verified [136; 175]. However, the number of discovered miRNAs is increasing [65], and each miRNA is thought to regulate a few hundred targeted genes in mammals [176]. Therefore, the problem of finding genes that are regulated by miRNAs is of a great importance to a better understanding of biological processes. Identifying experimentally miRNA targets is a laborious process with time-consuming and expensive experiments, and therefore computational methods for miRNA target prediction are applied for narrowing down potential candidates for experimental validation. The first computational target prediction tools were published and made available in 2003 [85; 115], and over the past decade a variety of computational methods has been developed for identifying putative targets of miRNAs. Besides other features, accessibility, conservation and thermodynamics are important factors for miRNA-target interaction.

It has been shown that the regulation of gene expression by miRNAs is a complex process and that dysregulation of miRNA networks has been implicated in several diseases. Different regulatory mechanisms can control miRNA expression and cause its alteration in human diseases [177]. miRNA expression can be modulated as a consequence of defects in the miRNA biogenesis machinery. A deregulation of miRNA expression can also be a result of increased or decreased transcription due to an altered transcription factor activity [177]. Another reason for disturbed miRNA-mediated gene regulation are polymorphisms within miRNA-binding target sites, commonly called miRSNPs.

The impact of SNPs on minimum free energy mRNA conformations has been comprehensively studied in [178]. The authors analyzed a total number of 34 557 SNPs in 12 450 genes. The minimum free energy conformations were calculated by using `RNAfold`. The authors provide a great variety of data about the distribution of SNPs within mRNA transcripts and their effect on minimum free energy values

of secondary structures as well as on the profile of the ensemble of suboptimal structures and structures with high Boltzmann probabilities, see also [36; 96; 179] for the impact of SNPs on the ensemble of secondary structures. The study [178] also includes the analysis of various subsets of transcripts and SNPs, e.g. with respect to the transcript length and different SNP types.

Martin *et al.* [179] study structural changes induced by SNPs in the 5'UTR of the human FTL gene in conjunction with associated SNPs that restore the overall wild-type ensemble of secondary structures, thus leading to the notion of structure-stabilizing haplotypes. The authors also analyzed the stabilizing effect of multiple structure-stabilizing haplotypes on binding sites of miRNAs and RNA binding proteins (nine cases of 3'UTRs and one case of 5'UTR). As pointed out by the authors, the findings suggest that certain SNP pairs are conserved in the human population because they stabilize ensembles of mRNA conformations.

In [180], the authors used SNPs cases to either cause a destabilization of the interaction due to changes in the free binding energy or a change in target accessibility due to alterations in the RNA secondary/tertiary structure. Their results suggest to not only consider the influence of a SNP on the miRNA target interaction through introduced mismatches but also on miRNA target accessibility through secondary structure alterations caused by SNPs inside or in the vicinity of miRNA target sites (miRTSs) when searching for a possible disease association.

Consequently, the aim of our work is to analyze the impact of SNPs on the thermodynamics of the mRNA-miRNA interaction and accessibility of the target and the role of computational methods in predicting this impact. To this end, we researched the recent literature for experimental studies on the impact of SNPs on miRNA bindings. We carried out a keyword search in the `PubMed` database using the search terms "miRNA", "target", "polymorphism/variant", "disease" and "3'UTR". Moreover, some of the publications were sourced from the Human microRNA Disease Database (`HMDD`) [181]. All miRSNP-disease association studies were included in the present work if they met the following three conditions:

- (i) The expression levels of both alleles involved are analyzed experimentally by SNP genotyping (for some instances, a combination of clinical association studies and strong *in silico* results) or related methods involving PCR

experiments.

- (ii) The underlying allele information can be extracted as consistent data from the NCBI database, the dbSNP database and the Ensembl database (3'UTR transcripts from BioMart).
- (iii) Identifying all metastable conformations within an energy offset δE above MFE conformations is computationally feasible (3'UTR length $\lesssim$ 1100 nt).

Meeting all three conditions at least to a certain extent reduced the number of case studies we found to a set of 34 RNA instances.

The sequence IDs were retrieved the NCBI Single Nucleotide Polymorphism Database (dbSNP) of nucleotide sequence variation. We also utilized miRdSNP [170] and mirTarbase [136] for retrieving information related to wild type and variant alleles, ensuring this way a maximum consistency between the publication and the different databases. A detailed description of all instances in terms of Ensembl ID codes is provided in Supplementary Material in Section .1, where for each case, we explain how the UTR is retrieved from the relevant publication and how the differences in the UTR annotation are resolved. The sequence length refers to data directly obtained from the NCBI database together with transcript information provided by the Ensembl Biomart database, and the length ranges between 124 nt and 1177 nt.

Subsequently, we analyzed the corresponding publications to retrieve information about the common followed framework and more importantly the *in-silico* analysis carried out in most of these studies.

2.2 Disease-associated miRSNPs Dataset

The selection of our dataset was governed by the need of having miRNA-mRNA interactions with a high level of experimental validation. We extracted 34 instances of [mRNA/3'UTR;SNP;miRNA] interactions, from comprehensive studies based upon PCR and/or luciferase reporter assays, considered to be the best evidence of direct regulation. The data are sourced from the following publications:

[LIG3;rs4796030;miR-221]: The case is analyzed in [182]. The SNP rs4796030 is defined by A↔C at 3'UTR position 83 of NM_002311.4. The 3'UTR length is 124 nt. The authors studied bladder cancer cases by using genotyping assays. The authors conjecture a stronger inhibitory effect of miR-221 for the C-allele.

[MSLN;rs1057147;miR-611]: The case is analyzed in [183]. The SNP rs1057147 is defined by G↔A at 3'UTR position 69 of NM_001177355. The 3'UTR length is 132 nt. Dual-luciferase assays showed a significantly lower reporter activity when the vector harbored the G allele as compared to the A allele. miR-611 mimic caused a reduced reporter activity of vectors harboring the G allele, concluding a stronger inhibitory effect of miR-611 for the G allele.

[CGA;rs6631;miR-1302]: The case is analyzed in [184]. The SNP rs6631 is defined by T↔A at 3'UTR position 215 of NM_000735. The 3'UTR length is 258 nt. Results of dual-luciferase reporter assays reveal that miR-1302 negatively regulates CGA, and the substitution of T by A at rs6631 within the binding site disrupts its regulation. The authors observe a stronger inhibitory effect of miR-1302 for the A-allele.

[CBR1;rs9024;miR-574-5p]: The case is analyzed in [185]. The SNP rs9024 is defined by G↔A at 3'UTR position 133 of NM_001757.2. The 3'UTR length is 284 nt. The authors employed dual-luciferase assays to evaluate the miRNA binding to both alleles and observed a stronger inhibitory effect of miR-574-5p for the A-allele.

[HTR3E;rs56109847;miR-510-5p]: The case is analyzed in [186], see also [174]. The SNP rs56109847 (rs62625044) is defined by G↔A at 3'UTR position 76 of NM_001256614.1. The 3'UTR length is 302 nt. The authors measure luciferase activity to evaluate the miRNA binding to both alleles and observe a stronger inhibitory effect of miR-510-5p for the G-allele, see Figure 1B in [186].

[KRT81;rs3660;miR-17]: **[KRT81;rs3660;miR-20b]**: The case is analyzed in [187]. The SNP rs3660 is defined by G↔C at 3'UTR position 102 of NM_002281. The 3'UTR length is 342 nt. An increased expression of the reporter gene for the C allele of rs3660 compared with the G allele was observed by luciferase assay. The G to C change reduced binding efficiency of miR-20b and miR-17-5p to KRT81 mRNA, leading to decreased translational repression, thereby increased

KRT81 expression in rs3660C allele compared with rs3660G allele.

[**SPI1;rs1057233;miR-569**]: The case is analyzed in [188]. The SNP rs1057233 is defined by C↔T at 3'UTR position 330 of NM_003120. The 3'UTR length is 369 nt. The authors investigate a possible association between SPI1 polymorphisms and systemic lupus erythematosus(SLE). Transfection experiments demonstrated that miR-569 can inhibit the expression of a reporter construct containing the non-risk allele (C), but not the risk allele (T), suggesting that the loss of microRNA-mediated regulation of expression could be a potential molecular mechanism of overexpression of SPI1 that might contribute to the occurrence of SLE.

[**HLA-G;rs1063320;miR-148a-3p**]: The case is analyzed in [189]. The SNP rs1063320 is defined by C↔G at 3'UTR position 233 of NM_002127.5. The 3'UTR length is 386 nt. The study aims at exploring factors affecting the asthma risk. The authors used real-time PCR and luciferase assays for measuring expression levels of miRNAs and C/G-alleles and found evidence for a stronger inhibitory effect of miR-148a-3p for the G-allele.

[**MTHFD1L;rs7646;miR-197**]: The case is analyzed in [190]. The SNP rs7646 is defined by A↔G at 3'UTR position 120 of NM_015440. The 3'UTR length is 393 nt. The authors study polymorphisms in the mitochondrial folate gene MTHFD1L associated with the risk of neural tube defects (NTDs). Their results indicate that miR-9 and miR-197 specifically downregulate MTHFD1L levels in HEK293 and MCF-7 cells and that SNP rs7646 significantly affects miR-197 binding affinity to the MTHFD1L 3'UTR, causing more efficient post-transcriptional gene repression in the presence of the allele that is associated with increased risk of NTDs. These findings were verified *in vitro* using the luciferase reporter assays.

[**NFKBIA;rs696;miR-449a**]: The case is analyzed in [191]. The SNP rs696 is defined by G↔A at 3'UTR position 126 of NM_020529. The 3'UTR length is 502 nt. Experiments showed that miR-449a reduced the relative luciferase activities via the NFkBIA 3'UTR target site created by the A allele. The results indicate that A allele strengthens the binding of miR-449a with 3'UTR of NFkBIA, which in turn inhibits the expression of NFkBIA.

[**TYMS;rs2790;miR-1248**]: The case is analyzed in [192]. The SNP rs2790

is defined by A↔G at 3'UTR position 89 of NM_001071. The 3'UTR length is 502 nt. *In vitro* functional study for TYMS rs2790 was carried out. Luciferase assays showed a lower expression level for rs2790 G allele as compared with A allele, and the hsa-miR-1248 had an effect on modulation of TYMS gene. The authors conclude a stronger inhibitory effect of miR-1248 for the G allele.

[CCNE1;rs3218073;miR-151a]: The case is analyzed in [193]. The SNP rs3218073 is defined by C↔T at 3'UTR position 332 of NM_001238. The 3'UTR length is 531 nt. Luciferase reporter assays were used to indicate that CCNE1 was a direct target of miR-151, and the rs3218073 T to C change resulted in altered regulation of CCNE1 expression. These findings support the hypothesis that miR-151 directly targets CCNE1 expression, and the variant T allele alters the dependence of CCNE1 3'UTR, thus allowing increased CCNE1 expression in the presence of this variant allele.

[NPM1;rs34351976;miR-337]: The case is analyzed in [194]. The SNP rs34351976 is defined by a delT mutation at 3'UTR position 165 of NM_002520. The 3'UTR length is 572 nt. Bioinformatics analysis combined with luciferase reporter assay demonstrated that the delT polymorphism creates an illegitimate binding site for miR-337-5p, which results in decreased protein levels.

[AGT;rs7079;miR-31]: **[AGT;rs7079;miR-584]**: The case is analyzed in [195]. The SNP rs7079 is defined by C↔A at 3'UTR position 556 of NM_000029. The 3'UTR length is 618 nt. For AGT gene, results of the experiments showed that miR-31 and miR-584 bind to the C allele of the hAGT gene and downregulate the luciferase gene expression either in human kidney or human liver cells.

[NCSTN;rs141849450;miR-455]: The case is analyzed in [196]. The SNP rs141849450 is defined by delCA at 3'UTR position 515 of NM_001290184. The 3'UTR length is 681 nt. Using luciferase-based assays, the authors demonstrated that rs113810300 and rs141849450 SNPs affected miRNA-mediated repression of Nicastrin. Notably, rs141849450 completely abolished the miR-455-mediated repression of Nicastrin. On the other hand, the seed region SNP delCA515-516 reduced miR-455-mediated repression. The authors observe a stronger inhibitory effect of miR-455 for the wild-type allele.

[SMUG1;rs2233921;miR-770]: The case is analyzed in [197]. The SNP

rs2233921 is defined by G↔T at 3'UTR position 80 of NM_001243787. The 3'UTR length is 628 nt. An *in vitro* assay was used to investigate whether rs2233921 alleles were associated with differential gene expression. The G to T point mutation resulted in a reduction of the expression of luciferase in all the experiments performed. Although not a particular miRNA was selected, miR-770 was selected as a candidate miRNA suggested by a review paper citing the work [198].

[**PARP1;rs8679;miR-145-5p**]: The case is analyzed in [182]. The SNP rs8679 is defined by T↔C at 3'UTR position 607 of NM_001618.3. The 3'UTR length is 769 nt. The methodology is the same as for [LIG3;rs4796030;miR-221], and the authors presume an additive effect of both instances on bladder cancer risk. For PARP1/rs8679, the authors conjecture a stronger inhibitory effect of miR-145-5p for the T-allele.

[**WFS1;rs1046322;miR-668-3p**]: The case is analyzed in [199]. The SNP rs1046322 is defined by G↔A at 3'UTR position 253 of NM_001145853.1. The 3'UTR length is 779 nt. Expression levels of luciferase assays are measured for both alleles, with a stronger inhibitory effect of miR-668-3p for the G-allele.

[**MSX1;rs12532;miR-3649**]: The case is analyzed in [200]. The SNP rs12532 is defined by A↔G at 3'UTR position 276 of NM_002448. The 3'UTR length is 790 nt. The sequence analysis indicated that SNP rs12532 might alter the binding ability of miR-3649, confirmed by luciferase activity assay showing a lower expression level of rs12532 A allele compared with that of the G allele. The authors conclude a stronger inhibitory effect of miR-3649 for the A allele.

[**EFNA1;rs12904;miR-200c**]: The case is analyzed in [201]. The SNP rs12904 is defined by G↔A at 3'UTR position 154 of NM_182685. The 3'UTR length is 843 nt. The authors investigated miRNA binding sites in SNPs and whether they interfere with gastric cancer (GC) susceptibility. Luciferase assays indicated EFNA1 as the target of hsa-miR-200c and rs12904 G→A change resulted in altered regulation of luciferase expression.

[**IL-23R;rs10889677;let-7e**]: The case is analyzed in [202]. The SNP rs10889677 is defined by C↔A at 3'UTR position 309 of NM_144701.2. The 3'UTR length is 851 nt. The authors study risk factors for inflammatory bowel diseases where

they employ real-time PCR and luciferase assays for measuring expression levels, and finally conclude a stronger inhibitory effect of let-7e for the C-allele. [IL-23R;rs10889677] is also associated in [203] with breast cancer development.

[**RYR3;rs1044129;miR-367**]: The case is analyzed in [204]. The SNP rs1044129 is defined by A↔G at 3'UTR position 839 of NM_001036.3. The 3'UTR length is 880 nt. The authors study risk factors of breast cancer development. The authors employ real-time PCR and luciferase assays for measuring expression levels. The authors observe a stronger inhibitory effect of miR-367 for the A-allele.

[**AGTR1;rs5186;miR-155-5p**]: The case is analyzed in [180]. The SNP rs5186 is defined by A↔C at 3'UTR position 86 of NM_032049.3. The 3'UTR length is 888 nt. The authors analyze luciferase assays for measuring expression levels and observe a stronger inhibitory effect of miR-155-5p for the A-allele, see 'long case' in Figure 2C in [180].

[**FGF20;rs12720208;miR-433-3p**]: The case is analyzed in [158]. The SNP rs12720208 is defined by C↔T at 3'UTR position 182 of NM_019851.2. The 3'UTR length is 903 nt. The authors analyze luciferase assays for measuring expression levels and observe a stronger inhibitory effect of miR-433-3p for the C-allele. We note that in this specific case we used the submission ss20399075 instead of the default `dbSNP` entry ss28476621 for consistency with the `NCBI` entry of NM_019851.2 and the corresponding 3'UTR transcript entry ENST00000180166 at `Ensembl`.

[**DROSHA;rs10719;miR-27b**]: The case is analyzed in [205]. The SNP rs10719 is defined by T↔C at 3'UTR position 92 of NM_013235. The 3'UTR length is 937 nt. Luciferase reported gene assay confirmed that rs10719 T to G substitution disrupted the binding site for hsa-miR-27b, resulting in increased levels of DROSHA protein. The authors observe a stronger inhibitory effect of miR-27b for the T-allele.

[**HOXB5;rs9299;miR-7-5p**]: The case is analyzed in [206]. The SNP rs9299 is defined by G↔A at 3'UTR position 141 of NM_002147.3. The 3'UTR length is 952 nt. The miRNA-3'UTR binding is studied in the context of bladder cancer development. Both real-time PCR and luciferase reporter assays are applied gene expression measurements. The authors observe a stronger inhibitory effect

of miR-7-5p for the A-allele.

[RAD51;rs7180135;miR-197-3p]: The case is analyzed in [182]. The SNP rs7180135 is defined by G↔A at 3'UTR position 718 of NM_002875.4. The 3'UTR length is 978 nt. The methodology is the same as for [LIG3;rs4796030;miR-221] and [PARP1;rs8679;miR-145-5p]. The authors conjecture a stronger inhibitory effect of miR-197-3p for the G-allele.

[REV3L;rs465646;miR-25]: The case is analyzed in [207]. The SNP rs465646 is defined by C↔T at 3'UTR position 461 of NM_002912. The 3'UTR length is 985 nt. The authors investigate association between REV3L polymorphisms and lung cancer risk. One of the strongest associations observed was for the 3'UTR C↔T polymorphism (rs465646). Surface plasmon resonance analysis and luciferase assays showed that the T allele demonstrated a stronger binding affinity for miR-25, resulting in significantly weaker reporter expression levels.

[ORAI1;rs76753792;miR-519a-3p]: The case is analyzed in [208]. The SNP rs76753792 is defined by C↔T at 3'UTR position 86 of NM_032790.3. The 3'UTR length is 1034 nt. The authors study the susceptibility of atopic dermatitis in Japanese and Taiwanese populations. Among other methods, real-time PCR is applied to gene expression analysis. The authors mention the impact of miRNAs as subject of future research, i.e. no specific miRNA is identified. Based upon miRNA target predictions for ORAI1 and the SNP position we selected miR-519a-3p for the present study. For miR-519a-3p, the binding prediction returned by `StarMir` is stronger for the C-allele.

[RAP1A;rs6573;miR-196a]: The case is analyzed in [209]. The SNP rs6573 is defined by A↔C at 3'UTR position 366 of NM_002884.2. The 3'UTR length is 1078 nt. The authors study how rs6573 affects the risk of esophageal squamous cell carcinoma. The regulatory function of miR-196a is analyzed by luciferase reporter assays. The authors conclude a stronger inhibitory effect of miR-196a for the A-allele.

[APP;T171C;miR-147]: The case is analyzed in [210]. The SNP T171C is defined by T↔C at 3'UTR position 171 of NM_000484. The 3'UTR length is 1120 nt. The authors provide proof-of-principle that APP 3'UTR polymorphisms could affect Alzheimer's disease risk through modulation of APP expression reg-

ulation. Using luciferase-based assays, the authors could show that the T171C variant inhibited miR-147 binding.

[**IL1A;rs3783553;miR-122**]: The case is analyzed in [211]. The SNP rs3783553 is defined by a TTCA insertion at 3'UTR position 928 of NM_000575. The 3'UTR length is 1152 nt. It was shown *in vitro* and *in vivo* that the variant allele consisting of a 4-bp (TTCA) insertion disrupts the binding sites for miR-122, thereby increasing the expression of IL-1A. Therefore the authors conclude a stronger inhibitory effect of the studied miRNA for the wild type allele.

[**CD133;rs2240688;miR-135b**]: The case is analyzed in [212]. The SNP rs2240688 is defined by A↔C at 3'UTR position 667 of NM_0011458. The 3'UTR length is 1167 nt. The authors investigated the associations between putative functional SNPs in CD133 and both the risk and the survival of lung cancer in southern Chinese. Functional assays including Real-time PCR and luciferase-based assays revealed that the A to C transition gained a new binding of the microRNA hsa-miR-135b and decreased the CD133 expression.

[**PDCD1;rs10204525;miR-4717**]: The case is analyzed in [213]. The SNP rs10204525 is defined by G↔A at 3'UTR position 889 of NM_005018. The 3'UTR length is 1177 nt. In this study, three miRNAs were predicted to putatively interact with PD1 rs10204525 polymorphic site of allele G. One of them, miRNA-4717, was demonstrated to allele-specifically affect luciferase activity in a dose-dependent manner in cells transfected with vectors containing different rs10204525 alleles. The authors conclude a stronger inhibitory effect of miR-4717 for the G allele.

A summary of the dataset is given in Table 2.1.

2.3 Dataset Analysis

2.3.1 Common Framework for miRSNPs-disease Association Studies

We extracted the common working scheme followed in most of the comprehensive studies of miRSNPs described in the previous section. A flowchart summarizing

the common steps is given in Figure 2.1. The first step is the selection of candidate genes associated with a selected disease, its risk and prognosis. Usually, selected genes are confirmed to be significantly associated with the selected disease and are retrieved from one or more previous association studies between gene polymorphisms and disease risk. The selected genes are narrowed down to retain only polymorphic 3'UTRs where SNPs were, *in-silico*, able to affect the binding with specific miRNAs that are expressed in the relevant cell type. To this end, miRNA target prediction software are essentially applied to search for the putative SNPs within miRNA binding sites in the 3'UTR of candidate genes. Moreover, databases integrating data on miRNA expression accross different tissues and cell types are used. The selected SNPs are usually assessed in appropriate case-control studies, where allele and genotype frequencies are statistically analyzed. The next stage usually consists of experimentally verifying whether the predicted 3'UTRs are truly target sites for the predicted miRNAs and whether the SNPs affect the strength of the binding. Experiments are carried out on cell lines transfected with the Renilla/Luciferase gene reporters, where the 3'UTRs of the Luciferase is replaced by the 3'UTRs of the two genes (either in their common wild-type or variant forms). Synthetic miRNAs are also transfected in order to measure the Luciferase/Renilla ratio under different miRNA binding, to detect whether the polymorphism truly affect the binding to the 3'UTR. This last step allows reaching the conclusion if the selected miRNA binds differentially to the wild-type and the variant 3'UTRs and therefore should help to elucidate functional consequences of the candidate miRSNPs in the selected disease.

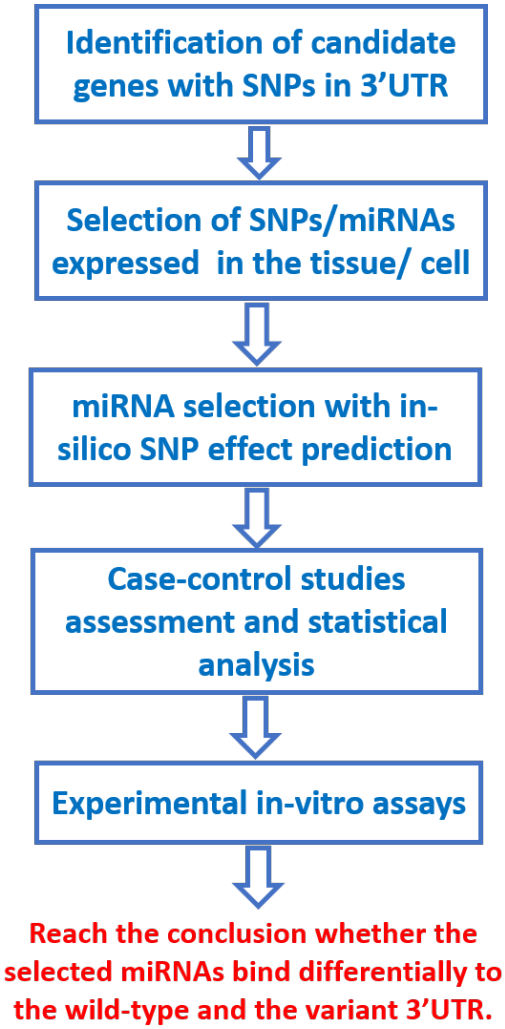


Figure 2.1: The common workflow followed in most miRSNP-disease association studies.

Table 2.1: Summary of the disease associated SNPs dataset used in this study. L(3'UTR) nt indicates the length of the relevant 3'UTR. The average length is 703. W-allele (S-allele) represents the allele with the weaker (stronger) miRNA inhibitory effect. Whether the allele represents the wild-type (w) or the variant allele (v) is indicated in brackets. The SNP ID, NCBI's dbSNP official SNP identifiers ("rs-numbers") and the position within the 3'UTR are given.

	L(3'UTR) nt	W-allele	S-allele	miRNA	SNP pos	SNP ID	Year	Disease
LIG3	124	A(w)	C(v)	221	83	rs4796030	2012	Bladder cancer
MSLN	132	A(v)	G(w)	611	69	rs1057147	2014	Malignant pleural mesothelioma
CGA	258	T(w)	A(v)	1302	215	rs6631	2011	Infertility
CBR1	284	G(w)	A(v)	574	133	rs9024	2012	Liver and heart disease
HTR3E	302	A(v)	G(w)	510	76	rs56109847	2008	Diarrhea predominant irritable bowel syndrome
KRT81	342	C(v)	G(w)	17/20b	102	rs3660	2015	Lung cancer
SPI1	369	T(v)	C(w)	569	330	rs1057233	2011	Systemic lupus erythematosus
HLA-G	386	C(w)	G(v)	148a	233	rs1063320	2007	Asthma
MTHFD1L	393	A(w)	G(v)	197	120	rs7646	2014	Neural tube defects
NFKBIA	502	G(w)	A(v)	449a	126	rs696	2011	Colorectal cancer
TYMS	502	A(w)	G(v)	1248	89	rs2790	2014	Lung cancer
CCNE1	531	T(v)	C(w)	151a	332	rs3218073	2013	Nasopharyngeal carcinoma
NPM1	572	T(w)	delT(v)	337	165	rs34351976	2013	leukemia
AGT	618	A(v)	C(w)	31/584	556	rs7079	2013	Hypertension
NCSTN	623	delCA(v)	CA(w)	455	515	rs141849450	2014	Alzheimers disease
SMUG1	628	G(w)	T(v)	770	80	rs2233921	2013	Colorectal cancer
PARP1	769	C(v)	T(w)	145	607	rs8679	2012	Bladder cancer
WFS1	779	A(v)	G(w)	668	253	rs1046322	2013	Wolfram syndrome
MSX1	790	G(v)	A(w)	3649	276	rs12532	2014	Non-syndromic orofacial clefts
EFNA1	843	A(v)	G(w)	200c	154	rs12904	2014	Gastric cancer
IL-23R	851	A(v)	C(w)	let-7e	309	rs10889677	2012	Inflammatory bowel disease
RYR3	880	G(v)	A(w)	367	839	rs1044129	2011	Breast cancer
AGTR1	888	C(v)	A(w)	155	86	rs5186	2012	Coronary artery disease
FGF20	903	T(v)	C(w)	433	182	rs12720208	2008	Parkinson disease
DROSHA	937	C(v)	T(w)	27b	92	rs10719	2013	Bladder cancer
HOXB5	952	G(w)	A(v)	7	141	rs9299	2012	Bladder cancer
RAD51	978	A(v)	G(w)	197	718	rs7180135	2012	Bladder cancer
REV3L	985	C(w)	T(v)	25	460	rs465646	2012	Lung cancer
ORAI1	1034	T(v)	C(w)	519a	86	rs76753792	2012	Dermatitis
RAP1A	1078	C(v)	A(w)	196a	366	rs6573	2012	Esophageal squamous cell carcinoma
APP	1120	C(v)	T(w)	147	171	T171C	2011	Alzheimers disease
IL1A	1152	TTCA(v)	-(w)	122	928	rs3783553	2009	Hepatocellular carcinoma
CD133	1167	A(w)	C(v)	135b	667	rs2240688	2013	Lung cancer
PDCD1	1177	A(v)	G(w)	4717	889	rs10204525	2015	HBV-associated liver disease

2.3.2 Commonly Used microRNA Target Prediction Methods

A crucial step in the process of miRSNP-disease association is the selection of candidate SNPs, giving their predicted impact on binding with a selected set of miRNAs. To this end, an *in-silico* analysis involving one or more miRNA target prediction methods needs to be carried out. Excluding the case of ORAI1, where no particular miRNAs were mentioned in the work, we had a look at this particular stage in all other 33 publications and listed the most frequently used miRNA target prediction tools. Apart from the FGF20 case, where no particular computational method was mentioned, 23 out of the other 32 articles used `TargetScan` web server directly or via the `PolymiRTS` database. `miRanda` came second with 19 publications mentioning the use of this algorithm directly or via online databases using `miRanda` such as `SNPinfo` [214] and `MirSNP` [215]. It was followed by `MicroCosm-miRBase` used in 13 instances, `PicTar` in 10 publications, `DIANA-microT` in 6, `MicroInspector` in 5, `RNAhybrid` in 4 and finally `PITA` in only 3 publications. Moreover, in 22 out of the 32 works, the authors selected miRSNPs with the highest support from the available tools, using the predictions of several algorithms. More importantly, for the cases where variant alleles are expected to have a stronger inhibitory effect, tools which allow sequence entry may be favored; 12 instances belong to this category and all these cases, with the exceptions of NFKBIA and REV3L instances, used either `miRanda`, `PITA` or `RNAhybrid` in their miRNA target predictions stage. As an example `miRanda` was used in the case of LIG3 [182] to compute the difference in ΔG for the two alleles (wild-type allele ΔG minus variant allele ΔG) calculated as $\Delta\Delta G$. Looking closely at the *in-silico* analysis in the case of NFKBIA, the predicted target site given in [191] (see Figure 2) covers the positions 110-124 of the NFKBIA 3'UTR and therefore does not include the SNP position 126, as given by the `dbSNP` database. The authors searched for target microRNAs of NFKBIA using the prediction tools `TargetScan` and `MicroCosm`. Thus, we had a look at the predictions given by these two tools and could not find a relevant binding site with the SNP position inside. In the case of REV3L, target prediction tools `Targetscan` and `Pictar` both predicted that the REV3L 3'UTR harbours two binding sites for

miR-25/32/92/363/367 and the SNP is located within one of these binding sites. Although the precise binding site was not given in the publication, we looked at the `TargetScan` and `Pictar` predictions and failed to find any binding site with the SNP position 460 inside.

2.4 Dataset Coverage by Online SNP Effects Prediction Databases

Recently, several databases studying disease associated SNP effects on miRNA bindings have been released. These databases follow similar algorithms as those used in miRNA target prediction tools in order to detect the effects of the SNPs on miRNA bindings. These algorithms are either run on the whole genome and then results are stored in a database, or come as web-based applications where the SNP effect is computed in real time. Usually, the user can query these databases using SNP, gene or miRNA IDs. A number of these databases offer valuable repositories for experimentally validated, disease associated, SNPs which have been verified to affect a miRNA binding. In this section, we compare results of some of the well known available databases for predicted SNP effects. Note that at the time of writing, the database `Patrocles` is not working. The latest release of the `miRBase` database is Release 21 and the Genome Reference Consortium most recent release of the human genome assembly is GRCh38. The latest release of `NCBI dbSNP` database is build 146. The case [APP;T171C;miR-147] has an unknown SNP ID and therefore was removed from this analysis.

2.4.1 MicroSNiPer

`MicroSNiPer` is a web-based application which predicts the impact of a SNP on putative miRNA targets. `MicroSNiPer` interrogates the 3'UTR and predicts if a SNP within a target site will disrupt/eliminate or enhance/create a miRNA binding site. This application computes these sites and examines the effects of SNPs in real time. Although `MicroSNiPer` current update is using previous gene annotation releases (human genome GRCh37/hg19 release and `dbSNP` build 137), users can input their own 3'UTR sequence and SNPs, enabling them to analyse

novel SNPs. However, there is less flexibility in the fact that users are unable to enter specific miRNAs and can only rely on the output display showing the miRNA-binding sites, using the old `miRBase` Release 19. Although `MicroSNiPer` gives a straightforward graphical representation of the binding sites, a numerical value, a score or an energy value is not provided to assess the binding. As an example, in the case of [REV3L;rs465646;miR-25] and [KRT81;rs3660;miR-17], both wild type and variant alleles have predicted binding sites with the same seed length (8 bp) and although the binding sites differ, it is not clear which allele is predicted to have a stronger binding with hsa-miR-25-3p and hsa-miR-17-5p respectively and therefore no conclusion can be drawn from this result. `MicroSNiPer` correctly predicted SNPs to affect the interactions between the studied miRNAs and the 3'UTR in 15 instances.

2.4.2 miRdSNP

`miRdSNP` is a database of manually curated dSNPs on the 3'UTRs of human genes from available publications in `PubMed`. `miRdSNP` annotates genes with experimentally confirmed targeting by miRNAs and indexes miRTSs predicted by `TargetScan` and `PicTar` as well as potential miRTSs newly generated by disease associated SNPs. The database was released in 2011 and no further updates were released. Only 2 of the 33 instances from our dataset are marked as disease associated SNPs. We also found 5 instances where the miRNA is listed as predicted targeting the gene, but not marked as experimentally verified interactions. Only one case ([HLA-G;rs1063320;miR-148a-3p]) is hosted as experimentally validated SNP effect associated to a disease.

2.4.3 PolymiRTS Database 3.0

The Polymorphism in microRNA Target Site (`PolymiRTS`) database aims to identify SNPs that affect miRNA targeting in human and mouse. The `PolymiRTS` database was created by scanning 3'UTRs of mRNAs in human and mouse for SNPs and INDELs in miRTSs. Then, the potential downstream effects of these polymorphisms on gene expression and higher-order phenotypes are identified. The `PolymiRTS` database also includes polymorphisms in target sites that have

been supported by a variety of experimental methods and polymorphisms in miRNA seed regions. The last significant update to the PolymiRTS database dates from 2013. Using the search interface of PolymiRTS, we found 11 instances from our dataset listed as predicted target sites with no experimental support. Only 2 cases are listed as disease associated SNPs. Five instances (SPI1, FGF20, AGTR1, HTR3E, HLA-G) are listed as miRNA:gene interactions supported by a low-throughput experiment (luciferase reporter assay or Western blot).

2.4.4 mrSNP

mrSNP [216] provides a web service for researchers working especially with RNA-Seq Data, to predict the impact of a SNP in a 3'UTR on miRNA binding. The software accepts input SNPs with the related information containing the organism, the assembly according to which the mapping is done, the chromosome on which the SNP is located, the position of the SNP in the given chromosome, and the SNP alleles. However, the available chromosomal locations in this database are based on a previous human genome build (GRCh37/hg19). The prediction method applied is adapted from the DIANA-microT prediction tool and the binding energy is calculated using RNAhybrid. mrSNP correctly identified 9 of the SNPs disrupting the binding with the associated miRNA.

2.4.5 Mirsnpscore

Mirsnpscore [217] is a computational tool that can help identifying SNPs associated with diseases, by focusing on SNPs affecting miRNA-regulation of genes. The tool predicts the effects of miRSNPs and uses linkage disequilibrium to map these miRNA-related variants to SNPs of interest in GWAS. The used sequences are sourced from the human genome assembly hg18 and miRNA sequences are sourced from miRBase release 13.0 and 16.0. Therefore, some of the SNP queries were not found in the database or the SNP information were incorrect. Mirsnpscore successfully identified 6 of the SNP-miRNA pairs from our dataset.

2.4.6 miRNASNP 2.0

miRNASNP [218] aims to provide a resource of the miRNA-related SNPs, which includes SNPs in pre-miRNAs of human and other species, and target gain and loss by SNPs in miRNA seed regions or 3'UTR of target mRNAs. The last release 2.0 dates from 2014 and is based on miRBase 19 and dbSNP 137. In this last release, validated miRNA-mRNA interactions supported by experimental methods were also integrated from several databases including DIANA-TarBase, miRecords, miRTarBase and miR2Disease. Using the module 'targets gain/loss by SNP in gene's 3'UTR', we used the SNP IDs to search the available pre-compiled data. miRNASNP correctly predicts the effect of 7 SNPs. Among these cases, 3 instances were marked as experimentally verified human miRNA targets (HTR3E, AGTR1 and EFNA1). In the case of KRT81, miRNASNP predictions are in favor of the weaker allele. Four other cases were also included in the predictions: FGF20, CD133, AGT and SMUG1. However, the 3'UTR sequences, for the wild-type and variant alleles from our dataset, do not seem to match the data from the dbSNP database.

2.4.7 Discussion

The dataset analysis in the different databases is summarized in Table 2.2. The MicroSNiPer and PolymiRTS databases successfully detect the SNPs to disturb a binding site with the associated miRNA in respectively 15 and 16 instances from the dataset, therefore performing better than the other platforms. Some databases restrict the user to catalogued 3'UTRs and SNPs, limiting the possibility for the user to enter a novel or unreported SNP. Only MicroSNiPer tool provides a relatively high degree of flexibility at the input stage by allowing the user to enter a user defined UTR sequence with an associated SNP. Furthermore, some of these available online resources provide only precomputed predictions, based on older version of human genome builds and/or previous releases of dbSNP and miRBase. Therefore the available predictions may be outdated and not relevant to the user. One limitation for most of these databases is the inability to study SNPs where polymorphisms are small insertions and deletions (INDELs) in miRTSs. These cases include NPM1, NCSTN and IL1A instances. Only

PolymiRTS handles this category of polymorphisms and therefore this tool was the only software to be able to successfully predict the SNP effect in the case of NCSTN.

Table 2.2: Dataset Coverage by Online SNP effect Prediction Databases.

	MicroSNiPer	miRdSNP	PolymiRTS	mrSNP	Mirsnpscore	miRNASNP
LIG3						
MSLN						
CGA						
CBR1						
HTR3E	✓		✓	✓		✓
KRT81	✓	✓	✓	✓		
SPI1	✓		✓			
HLA-G	✓	✓	✓	✓	✓	✓
MTHFD1L	✓				✓	
NFKBIA						
TYMS				✓		✓
CCNE1	✓		✓			✓
NPM1						
AGT	✓		✓			
NCSTN			✓			
SMUG1			✓			
PARP1						
WFS1	✓		✓	✓	✓	✓
MSX1	✓		✓	✓	✓	
EFNA1	✓	✓	✓	✓		✓
IL-23R						
RYR3						
AGTR1	✓		✓	✓		✓
FGF20	✓		✓			
DROSHA		✓				
HOXB5		✓				
RAD51	✓		✓	✓	✓	
REV3L		✓				
ORAI1						
RAP1A					✓	
IL1A						
CD133	✓		✓			
PDCD1	✓		✓			

2.5 Dataset Coverage by Experimentally Validated miRNA Target Interactions Databases

Our dataset represents 34 ,disease associated experimentally validated SNPs ,reported in the literature to affect miRNA binding and curated from low throughput experiments. We discuss in this section how much of this validated work

is recorded in the abundant number of databases, which host records of experimentally verified miRNA-gene interactions. Note that this section does not cover the actual effect of the SNPs on miRNA bindings. The SMUG1 and the ORAI1 cases were removed from the analysis, as the related publications do not mention a particular miRNA.

2.5.1 miRTarBase

`miRTarBase` is a curated database of miRNA-target interactions. The listed interactions are collected by manually surveying pertinent literature after data mining of the text systematically filter research articles related to functional studies of miRNAs. `miRTarBase` database was first released in 2010 and have been regularly updated. The latest release 6.0 dates from 2015. We searched for the miRNAs:gene interactions from our dataset and we were able to find 11 instances covered in this database.

2.5.2 DIANA-TarBase v7.0

`DIANA-TarBase` was initially released in 2006 and it was the first database aiming to catalog published experimentally validated miRNA:gene interactions.
`DIANA-TarBase v7.0` is the last update and was released in 2014. The user interface provides detailed information about positive or negative experimental results, the utilised experimental methodology, experimental conditions including cell/tissue type and treatment etc. We found that 10 out the 32 interactions from our dataset are covered in this database.

2.5.3 miRecords

`miRecords` is a resource for animal miRNA-target interactions. The Validated Targets component of `miRecords` is a large, high-quality database of experimentally validated miRNA targets resulting from meticulous literature curation. Although `miRecords` was last updated in 2013, only 3 out of the 32 interactions from our dataset are recorded.

2.5.4 miRWalk 2.0

The web-interface of `miRWalk2.0` [219] is classified into the Predicted Target (PTM) and the Validated Target (VTM) modules. The VTM module documents experimentally verified miRNA-target interaction information collected via an automated text-mining search and data from existing resources (`miRTarBase 4.0`, `PhenomiR 2.0`, `miR2Disease` 2008 and `HMDD 2.0`). This module was last updated in 2014. We used the validated gene-miRNA interaction information retrieval system to check and we found that 8 out of 32 interactions are covered.

2.5.5 HMDD v2.0

`HMDD` (the Human microRNA Disease Database) is a database that curated experiment-supported evidence for human miRNA and disease associations. The original `HMDD` database was released in 2007. However, the last update was released in 2013 and integrated experimentally verified disease-related miRNAtarget interactions. We were able to download the freely available text file containing a list of miRNA-disease association data from miRNA-target interactions and we could find 11 out of the 32 interactions from our dataset listed.

2.5.6 miR2Disease

`miR2Disease` is a manually curated database, which aims at providing a comprehensive resource of miRNA deregulation in various human diseases. Additionally, the database integrates manually extracted disease-related miRNA-target pairs, which were experimentally verified by luciferase reporter experiments. The database was created in 2008. However, there is no direct information about the last update release date. While looking at the experimentally verified miRNA-target file, available freely to download, we were able to find 4 out of the 32 publications from our dataset listed in `miR2Disease`.

2.5.7 Discussion

The dataset analysis in the different databases is summarized in Table 2.3. The HTR3E and the AGTR1 cases are covered by all databases. Fourteen of the 32 in-

stances from our dataset are not included in any database (LIG3, MSLN, CBR1, KRT81, MTHFD1L, NFKBIA, TYMS, NPM1,AGT, PARP1, WFS1, MSX1, APP and IL1A). `miRTarBase` and `HMDD` include equally the same number of publications from our study. However, they only have 6 publications in common. Some databases have a user submission form that allows users to notify the database curators of recent publications related to experimentally validated miRNA:gene interactions. These include `miR2Disease`, `HMDD` and `miRecords`. This can benefit to the database by having regular updates with new entries. While `DIANA-TarBase` contains more data related to each interaction, `miRTarBase` seemed to be more user friendly. As an example, while searching for the CD133 gene, the name was not recognised by `DIANA-TarBase 7.0`. However, in `miRTarBase`, the name was automatically converted to the commonly used Alias which is PROM1.

Table 2.3: Dataset Coverage by Experimentally Validated miRNA Target Interactions Databases.

	miRTarBase	DIANA-TarBase	miRecords	miRWalk	HMDD	miR2Disease
LIG3						
MSLN						
CGA					✓	
CBR1						
HTR3E	✓		✓	✓	✓	✓
KRT81						
SPI1	✓			✓	✓	
HLA-G	✓		✓	✓		✓
MTHFD1L						
NFKBIA						
TYMS						
CCNE1	✓		✓	✓	✓	
NPM1						
AGT						
NCSTN	✓					
PARP1						
WFS1						
MSX1						
EFNA1	✓		✓			
IL-23R					✓	
RYR3					✓	
AGTR1	✓		✓	✓	✓	✓
FGF20	✓		✓	✓	✓	✓
DROSHA			✓			
HOXB5	✓		✓		✓	✓
RAD51	✓		✓		✓	
REV3L	✓		✓		✓	
RAP1A					✓	
APP						
IL1A						
CD133					✓	
PDCD1						

2.6 Dataset Coverage by Online miRNA Target Prediction Databases

In 22 out of the 32 works used in this study, the authors considered that the combination of several miRNA-target prediction approaches would greatly reduce the possibility of false positives. There is also a strong demand among researchers for integrated miRNA target databases. In fact, several resources have been established to systematically retrieve precompiled lists of predicted targets, generated by multiple target prediction methods. In this section, we analyze some of the commonly used databases and check whether they predict the miRNAs:3'UTR interactions from our study. Of particular interest were the cases where the wild types are experimentally proven to interact with the miRNAs. Attention was particularly given to the sequences used for predictions. For consistency, the same `NCBI` Reference Sequence (RefSeq) accessions and miRNA IDs were used in the queries. Similar to the previous section, this section does not cover the actual effect of the SNPs on miRNA bindings.

2.6.1 miRecords

The Predicted Targets component of `miRecords` is an integration of predicted miRNA targets produced by 11 established miRNA target prediction programs. The targets predicted by the different tools were obtained by either submitting queries to the relevant web servers(`DIANA-microT`, `MicroInspector`, `miTarget`, `NBmiRTar` and `RNA22`), running a local implementation of the software (`miRanda` and `RNAhybrid`) or downloading the precompiled sets (`MirTarget2`, `PicTar`, `PITA` and `TargetScan/TargetScanS` Version 4.1). `miRecords` was last updated in 2013 and the sequences were based on older versions of `miRBase` and human genome assembly. If the user specifies the RefSeq accession or a gene name of an individual candidate target, the prediction is made in real time by submitting the request to each of the original predicting servers. In our case, we used the RefSeq IDs. However, some of the original servers were down (indicated by red symbols in the predictions table) and therefore the predictions were not complete for `MicroInspector`, `miTarget`, `NBmiRTar` and `RNA22` tools. The predictions

obtained for the 34 cases are summarized in Table 2.4. The miRNA was not found in the case of [PDCD1;miR-4717] and [MSX1;miR-3649]. The used RNA sequence could not be verified in the case of [WFS1;miR-668] and [HTR3E;miR-510]. [AGT;miR-31] is not predicted by any tool. The remaining cases were predicted by at least one tool.

Table 2.4: Dataset Coverage by miRecords database where predictions by 7 established miRNA target prediction programs are shown.

	DIANA-microT	MirTarget2	PicTar	PITA	miRanda	RNAhybrid	TargetScan 4.1
LIG3						✓	
MSLN				✓		✓	
CGA				✓		✓	
CBR1				✓			✓
HTR3E				✓		✓	✓
KRT81			✓	✓		✓	
SPI1				✓		✓	
HLA-G				✓		✓	
MTHFD1L				✓	✓	✓	
NFKBIA				✓	✓	✓	
TYMS				✓	✓	✓	
CCNE1				✓	✓	✓	
NPM1						✓	
AGT				✓	✓	✓	
NCSTN				✓		✓	
SMUG1						✓	
PARP1						✓	
WFS1						✓	✓
MSX1							
EFNA1		✓	✓	✓		✓	
IL-23R				✓	✓	✓	
RYR3				✓		✓	✓
AGTR1				✓	✓	✓	
FGF20				✓	✓	✓	
DROSHA				✓		✓	
HOXB5			✓	✓	✓	✓	
RAD51				✓	✓	✓	✓
REV3L		✓	✓	✓		✓	
ORAI1				✓	✓	✓	✓
RAP1A						✓	
APP				✓		✓	✓
IL1A				✓		✓	
CD133				✓	✓	✓	
PDCD1							

2.6.2 miRWalk2.0

The PTM module of `miRWalk2.0` database documents miRNA binding sites within the complete sequence of a gene and combines this information with a comparison of binding sites resulting from 13 existing miRNA-target prediction pro-

grams (DIANA-microTv4.0, DIANA-microT-CDS, miRanda-rel2010, miRBridge, miRDB4.0, miRmap, miRNAMap, doRiNA i.e.,PicTar2, PITA, RNA22v2, RNAhybrid2.1 and Targetscan6.2). All the possible targets are obtained from these established miRNA-target prediction programs, predicted without any threshold or filter, and stored in miRWalk database. PITA, miRanda, RNAhybrid and Targetscan were locally executed to identify miRNA binding sites within the putative target sequences. The generated miRNA binding sites were merged with the remaining 8 prediction datasets, gathered from the existing resources. miRBase release 20 and RefSeq release 61 are utilised as sources for the genomic sequences. Using the miRNA information retrieval system, we carried out miRNA-based searches by providing the relevant identifiers of miRNAs. Putative target genes tables containing precompiled predictions are then returned and used to search for the relevant RefSeq ID of the target gene. The predictions obtained for the 34 cases are summarized in Table 2.5. Apart from [NPM1;miR-337] and [MSLN;miR-611] instances, all other cases were successfully predicted by at least one miRNA-target prediction program.

2.6.3 starBase v2.0

starBase Pan-Cancer Analysis Platform is designed for deciphering Pan-Cancer Networks of lncRNAs, miRNAs, ceRNAs and RNA-binding proteins (RBPs) by mining clinical and expression profiles of 14 cancer types ($>$ 6000 samples) from The Cancer Genome Atlas (TCGA) Data Portal [138]. Additionally, starBase provides predicted miRNA-target interactions processed from five miRNA target prediction software programs (TargetScan, PicTar, PITA, miRanda and RNA22), last generated in 2013. Predicted miRNA-target interactions listed in the database are overlapped with CLIP-Seq data. The human genome sequences from UCSC hg19, miRNA IDs from miRBase Release 20 and the TargetScan Release 6.2 were used. In the query page of miRNA-mRNA interactions, users can enter a gene symbol and select one miRNA of interest to browse their relationships. Eight interactions from our dataset were predicted by at least one miRNA target prediction program. However, some of the other miRNAs were not listed in the available query list.

Table 2.5: Dataset Coverage by miRWalk where predictions by 11 established miRNA target prediction programs are shown. `miRBridge` program predicted only one case (AGT) and therefore was removed from the table.

	miRWalk	Microt4	miRanda	miRDB	miRMap	miRNAMap	Pictar2	PITA	RNA22	RNAhybrid	Targetscan
LIG3										✓	
MSLN											
CGA								✓		✓	
CBR1								✓		✓	
HTR3E	✓		✓			✓				✓	✓
KRT81	✓	✓	✓		✓			✓	✓	✓	✓
SPI1	✓		✓	✓	✓					✓	✓
HLA-G									✓		
MTHFD1L	✓				✓			✓		✓	✓
NFKBIA	✓		✓		✓	✓		✓	✓	✓	✓
TYMS	✓		✓	✓	✓			✓		✓	✓
CCNE1			✓		✓				✓	✓	✓
NPM1											
AGT			✓		✓			✓		✓	✓
NCSTN			✓		✓	✓		✓		✓	✓
SMUG1					✓					✓	
PARP1	✓									✓	
WFS1	✓		✓		✓					✓	✓
MSX1	✓	✓	✓		✓					✓	✓
EFNA1	✓		✓	✓	✓		✓			✓	✓
IL-23R	✓							✓	✓	✓	
RYR3		✓			✓		✓	✓		✓	✓
AGTR1			✓		✓					✓	✓
FGF20	✓		✓					✓		✓	✓
DROSHA		✓	✓		✓				✓	✓	✓
HOXB5		✓	✓		✓	✓	✓	✓		✓	✓
RAD51	✓		✓	✓	✓					✓	✓
REV3L	✓		✓	✓	✓	✓	✓	✓		✓	✓
ORAI1	✓	✓	✓					✓		✓	✓
RAP1A											
APP					✓					✓	
IL1A								✓		✓	
CD133										✓	
PDCD1	✓		✓							✓	✓

2.6.4 Discussion

Our dataset is a collection from literature of pairs of positive examples of miRNA-target genes. Thus, the lack of negative examples makes a direct comparison of prediction performance among tools, involving sensitivity and precision, not possible. However, a performance comparison in terms of the number of the experimentally validated interaction that are predicted by at least one miRNA target prediction tool can be done. Overall, `starBase` is the worst database in terms of predictions. The main reason for this performance is that although the 5 prediction algorithms are also used in `miRWalk`, only the conserved miRTSs, intersected with Ago CLIP clusters to gain CLIP-supported sites, are retrieved. Only 2 interactions were not predicted by `miRecords` (MSX1 and PDCD1) and 3 interactions were not predicted by `miRWalk` (MSLN, NPM1 and RAP1A). However, looking more closely at the predictions results, although 6 miRNA target prediction tools are used in both `miRecords` and `miRWalk`, the prediction results are different.

In `miRWalk` platform, it is stated that all the possible predicted targets are obtained from the established miRNA-target prediction programs with no threshold or filter. This information was not available in the case of `miRecords`. The default settings of some of miRNA target prediction tools, including `miRanda` and `RNAhybrid`, are set to ensure the maximal detection of targets. As a consequence, these tools are reported to have better sensitivity than other methods. However, this makes them prone to high false positive rates [220] leading to a lower specificity. In [99], the authors strongly recommend using optimized parameters instead of the default parameters. Therefore, the `RNAhybrid` tool is usually used to find the minimum free energy hybridization after filtering steps. In this study, we are focusing on measuring the sensitivity and not the specificity of the tools and therefore as expected, `RNAhybrid` performed the best in both platforms. The targets predicted by `RNAhybrid` were obtained by running a local implementation of the `RNAhybrid` algorithm provided by the authors (version 2.1 in `miRWalk` and version 2.2 in `miRecords`). Only 4 cases are not predicted by `RNAhybrid-miRWalk`. However, these cases are predicted by `miRecords-RNAhybrid`. Similarly, 3 cases are not predicted by `RNAhybrid-miRecords`, but these are predicted by `miRWalk-RNAhybrid`.

The targets predicted by miRanda were obtained by running a local implementation of the algorithm provided by the authors in both platforms. However, in miRecords Version 1.9 (2004) is indicated while miRanda 2010 version (3.3) is used in miRWalk. The miRanda tool from both databases predicts 9 cases in common. Three additional instances are predicted by miRecords-miRanda and 11 more cases are predicted by miRWalk-miRanda. In miRecords, the targets predicted by PITA version 6 from 2008 were downloaded from the PITA web site as a pre-compiled dataset. In miRWalk, an earlier 2007 version of the software was locally executed. PITA in miRWalk predicts 15 cases, all of them are also predicted by PITA in miRecords. Eleven other cases are predicted by PITA in miRecords only. The targets predicted by DIANA-microT were obtained by submitting queries to the DIANA-microT server in miRecords, where the 2004 version was used. DIANA-microT in miRecords does not predict any case. The prediction dataset of Diana-microT 4.0 and 5.0 from 2013 was downloaded and used by miRWalk. Six instances were predicted by Diana-microT-miRWalk. The targets predicted by TargetScan/TargetScanS Version 4.1 were downloaded from the tool web page as a pre-compiled dataset in miRecords. In miRwalk, it is indicated that TargetScan 6.1 program was locally executed with gene sequences RefSeq 61 and miRBase V20. Only 5 cases are predicted in common between the two platforms. Two are only predicted by TargetScan-miRecords while a further 17 cases are only predicted by TargetScan-miRWalk. PicTar in both databases predicts 4 cases. However, only 3 cases are in common. PicTar in miRecords predicts KRT81 while PicTar2 in miRWalk predicts RYR3. A 2007 version of PicTar precompiled set was used in miRecords while a precompiled set of doRiNA (PICTAR2) is used in miRWalk.

The difference in prediction results for the same tools can be a result of either using different genomic sequences, or using different threshold values of the different parameters, or finally a result of using different versions of the same miRNA-target prediction software. Although these databases are attractive for their ease of use, they need to be actively maintained to keep up with the regular updates of the annotation databases. Considering this limitation, we believe that the use of tools which allow sequence entry are more relevant to miRSNP-disease association studies.

2.7 Site Accessibility compared with the Conservation Feature in MicroRNA Target Prediction

While introducing our 34, manually curated, disease associated miRSNPs, we analyzed the relevant publications and we extracted the common framework followed in most of these studies. We also listed the most frequently used miRNA target prediction tools. While the conservation feature based tool `TargetScan` is the most widely used, structure-based miRNA target prediction models are hardly mentioned, even though, as discussed in Section 1.3.4, it has been shown that the accessibility criterion strongly influences the RNA hybridization process. We propose in this section to compare the performance of three individual miRNA target prediction methods based on ranked list of target predictions: `TargetScan` compared with the two structure-based methods `PITA` and `STarMir`.

`TargetScan` relies on the evolutionary conservation of target sites containing seed regions. `TargetScan` ranked lists are grouped by families, based on their identical seed sequence, with the idea that they should potentially have shared targets. Predicted targets of each miRNA family are sorted by total context+ score and the default ranked list does not show genes with only poorly conserved sites. The context+ score for a specific site is the sum of the contribution of these six features, calculated as in [107]: site-type contribution, 3' pairing contribution, local AU contribution, position contribution, TA (target site abundance) contribution and finally SPS (seed-pairing stability) contribution. For each predicted target of each miRNA, the sum of the context+ scores for the sites to that miRNA was calculated as the total context+ score. The predictions generated by `STarMir`, accessible from the the web searchable database `STarMirDB`, are based on an old human genome Build 36 version 3 dating from 2008 and similarly, the `PITA` catalog version 6 of predicted miRNA targets is based on miRBase version 11 and the human genome release hg18 from 2006. However, we have seen in the previous section that using different genomic sequences may lead to different prediction outcomes. Therefore, in order to use the same UTR and miRNA sequences, we proceeded with the following steps:

- Start with the `miRBase` database and search for the miRNA of interest. In the corresponding `miRBase` page, links to experimentally validated targets for the specified miRNA (`miRTarBase` and `DIANA-TarBase`) are given.
- Use `DIANA-TarBase` and `miRTarbase` databases to filter out validation methods with weaker evidence, to keep only experiments with Reporter assays, qPCR and/or Western Blot.
- Retrieve the ranked list of predicted targets by `TargetScan` for the specified miRNA.
- Locate the highest ranked in the list of predicted targets, that is an experimentally validated target.
- If this target exists and is in the top 10 of predicted targets, apply `PITA` and `STarMir` to the validated miRNA-target and then to the other ranked predicted targets in the top 10 of the `TargetScan` list. To this end, we used `STarMir` and `PITA` web servers, which allow users to provide mRNA and miRNA sequences, to generate predictions for the given miRNA with the same UTR and miRNA sequences used to generate the ranked list for `TargetScan`.
- Compare the `STarMir` scores and rank the targets for each miRNA. The same is applied to `PITA` scores.

After this process is finished, we obtain, for each miRNA, three ranked lists generated by the three tools of the same UTRs. In this section, we use a dataset comprising of more than 150 miRNAs. For each miRNA, we extract the top 10 target genes ranked by `TargetScan`. We only keep the miRNAs that have at least one experimentally confirmed target in the top 10, supported by strong experimental evidences (Reporter assay, Western blot or qPCR). This led to a reduced number of 36 analyzed miRNAs. We then compare the performance of the methods for each miRNA based on the rank of the first experimentally confirmed target. With respect to each miRNA, we score each method using a number called ranking score in the range of 1 to 3, with 1 indicating the best method and 3 the worst method. Finally, we calculate the ranking score of each

method by summing up its scores for all miRNAs. The lower the ranking score of a method, the better the method is. A summary of the rankings is provided in Supplementary Material in Section .2.

Using this scoring scheme, `PITA` has the best overall score while `STarMir` comes second and finally `TargetScan` comes last. This study shows that miRNA target prediction using structural information may outperform usual tools using the conservation feature.

2.8 Conclusion

In this chapter, we introduced the main motivation of our work on analyzing the disease associated SNP effect on miRNA bindings. We selected a dataset of 34 [mRNA/3'UTR; SNP; miRNA] instances from recent literature, based on strong expression level analyses and we analyzed the corresponding publications, concluding that the use of miRNA-target prediction methods is a crucial step in the process of miRSNP-disease association.

We subsequently wanted to investigate how online resources, made available to researchers for miRNA research, can be used in these miRSNPs-disease association studies. We presented a brief overview of the most widely used databases and repositories, mainly in three categories. The results obtained for the 34 cases are summarized in Table 2.6, where the instances are ordered in decreasing order of the total number of databases covering the case. Of particular interest were online databases available for predicting SNP effects on putative miRNA targets. These resources are of a great interest for researchers studying miRSNPs and we found that 12 instances out the 34 from the dataset were not predicted by any of the available repositories. We also found that very few resources allow a certain degree of flexibility at the input stage and most of the databases restrict the users to cataloged 3'UTRs, SNPs and miRNA sequences, which can be outdated and not relevant to the user. This represents the main limitation for users looking to predict the effects of novel or unreported SNPs.

Furthermore, we compared the results of two major computational approaches to miRNA target ranking prediction: conservation feature using `TargetScan` tool and target site accessibility feature using `PITA` and `STarMir` tools. We conclude

that using the site accessibility feature may improve miRNA-target predictions.

In our work, we plan to investigate how to improve detecting the effects of SNPs located in miRNA-3'UTR binding regions, on mRNA-miRNA bindings. We are particularly interested in the impact of these SNPs on miRNA bindings to metastable structures. Thus, as a first step, we will carry out an in-depth analysis of features related to the accessibility of binding regions in secondary structures representing relatively deep local minima, which applies to wild-type as well as to mutated sequences. If the findings of such an analysis are promising, we plan to devise a novel miRNA target prediction tool that can be used to evaluate SNP-effect on miRNA regulation.

Table 2.6: Summary of the dataset coverage indicating the number of online databases and repositories covering each case. SNP dbs refers to the number of SNP-effect prediction databases and Exp dbs indicates the number of the databases which offer repositories for experimentally validated miRNA-target interactions. The cases are ordered by decreasing number of Total which indicates the sum over the three studied categories of databases. Used Tools refers to the number of miRNA target prediction tools used in the *in-silico* analysis in the corresponding publication.

	L(3'UTR) nt	SNP dbs	Exp dbs	miRecords	miRWalk	starBase	Total	Used Tools
REV3L	985	1	3	4	9	5	22	2
EFNA1	843	5	2	4	7	2	20	2
HOXB5	952	1	4	4	8	3	20	4
HTR3E	302	4	6	3	5	0	18	1
AGTR1	888	4	6	3	4	0	17	1
RAD51	978	4	3	4	6	0	17	6
CCNE1	531	3	4	3	5	0	15	4
FGF20	903	2	5	3	5	0	15	NA
KRT81	342	4	0	3	8	0	15	2
HLA-G	386	6	4	2	1	1	14	5
SPI1	369	2	3	2	6	0	13	1
NFKBIA	502	0	0	3	8	1	12	2
TYMS	502	2	0	3	7	0	12	1
WFS1	779	5	0	2	5	0	12	2
NCSTN	623	1	1	2	6	1	11	1
DROSHA	937	1	1	2	6	1	11	4
MTHFD1L	393	2	0	3	5	0	10	2
AGT	618	2	0	3	5	0	10	2
MSX1	790	4	0	0	6	0	10	1
RYR3	880	0	1	3	6	0	10	3
ORAI1	1034	0	NA	4	6	0	10	NA
IL-23R	851	0	1	3	4	0	9	1
CD133	1167	2	1	3	1	0	7	1
PDCD1	1177	2	0	0	4	0	6	3
CGA	258	0	1	2	2	0	5	4
APP	1120	NA	0	3	2	0	5	3
IL1A	1152	0	0	2	2	0	4	1
CBR1	284	0	0	2	2	0	4	5
SMUG1	628	1	NA	1	2	0	4	1
PARP1	769	0	0	1	2	0	3	6
RAP1A	1078	1	1	1	0	0	3	4
LIG3	124	0	0	1	1	1	3	6
MSLN	132	0	0	2	0	0	2	5
NPM1	572	0	0	1	0	0	1	1

Chapter 3

Accessibility of microRNA binding sites in metastable RNA secondary structures in the presence of SNPs

In Chapter 2, we introduced the context of our work on how to improve detecting the effects of disease-associated SNPs on mRNA-miRNA bindings. We researched the recent literature for studies on the impact of some of these SNPs on miRNA bindings and we analyzed the corresponding publications. Moreover, we investigated the limitations of some of the available online resources for miRNA research. In this chapter, we study miRNA bindings to metastable RNA secondary structures close to minimum free energy (MFE) conformations in the context of SNPs and mRNA concentration levels, i.e. whether features of miRNA bindings to metastable conformations could provide additional information supporting the differences in expression levels of the two sequences defined by a SNP. In our study, the instances [mRNA/3'UTR; SNP; miRNA] were selected based on strong expression level analyses, SNP locations within binding regions and the

computationally feasible identification of metastable conformations. We identified 14 basic cases [mRNA; SNP; miRNA] of 3'UTR-lengths ranging from 124 up to 1078 nt reported in recent literature, and we analyzed the number, structure and miRNA binding to metastable conformations within an energy offset δE above MFE conformations. For each of the 14 instances, the miRNA binding characteristics are determined by the corresponding `STarMir` output. Among the different parameters we introduced and analyzed, we found that three of them, related to the average depth and average opening energy of metastable conformations, may provide supporting information for a stronger separation between miRNA bindings to the two alleles defined by a given SNP.

The work carried out in this chapter is a joint work between Ouala Abdelhadi Ep Souki, Luke Day, Andreas A. Albrecht and Kathleen Steinhöfel. All authors participated in the design of the methodology, the analysis of the results and the writing of the manuscript. Luke Day implemented the application `MSBind` and generated the data in Section .4 and Ouala Abdelhadi Ep Souki carried out the miRNA binding site and energy predictions from Section .3 and the analysis of the predictions. We note that this work has been published in [221].

3.1 Background

Within the past few years, analyzing concentration levels of microRNAs (miRNAs) and their putative messenger RNA (mRNA) targets has become a major topic in miRNA research.

Subkhankulova *et al.* [104] experimentally evaluated a parameterized analytical expression that estimates the number of genes g having t transcripts present in a single cell. The parameters are adjusted based on microarray data for a large number of genes extracted from single embryonic mouse neural stem cells, where the actual aim is the comparison of transcript numbers in phenotypically identical cells. The authors conclude from observed data for about 13,000 genes that the typical number of gene copies lies between 5 and 20, with 85% of genes having less than 100 copies in a single cell. Although the analysis is carried out for a specific cell type, the authors expect similar distribution results for a wide range of cell types. Arvey *et al.* [222] provide experimental evidence that

short RNAs (miRNAs and siRNAs) having a higher number of target transcripts within a single cell will downregulate each individual target gene to a lesser extent than those with a lower number of targets, which implies that the competition between target genes for a limited number of small RNAs may determine the degree of downregulation; see also Salmena *et al.* [223] for the concept of competing mRNAs associated with the number and distribution of multiple binding sites. Saito and Sætrom [224] provide data supporting the assumption that endogenous miRNAs preferentially target mRNAs with long 3'UTRs. The authors also discuss - among other features - the interaction between exogenous and endogenous miRNAs and critically assess the value of microarray data in the context of miRNA target prediction. The sequencing method developed by Mullokandov *et al.* [225] supports the assertion that only the most abundant miRNAs mediate target suppression. For example, deep sequencing of monocyte cells revealed the presence of about 310 miRNAs, with only about 40% of the miRNAs showing suppressing activity. For more than 80% of the targets, the corresponding miRNA was expressed above 100 reads per million. For the miRNA target prediction tool `TargetScan`, Garcia *et al.* [107] demonstrate how predictions may improve if target abundance is accounted for in binding scores. The impact of the life cycle of mRNAs on siRNA and microRNA efficacy is studied in [226]. The authors draw the conclusion that microRNA target prediction could be improved if data about mRNA turnover rates are incorporated into prediction tools. While [222; 226] focus on mRNA concentration levels, Cuccato *et al.* [227] propose different parameterised models of RNA interference that describe the effects of varying quantities of siRNAs. The models are derived from the basic equation $\mathrm{d}X_m/\mathrm{d}t = k_m - d_m X_m - \delta(X_m, X_s)$, where X_m and X_s are the mRNA and siRNA concentrations, k_m is the mRNA transcription rate, d_m the basal mRNA degradation rate, and $\delta(X_m, X_s)$ is the siRNA induced mRNA degradation rate. The models differ in the assumption about $\delta(X_m, X_s)$ and are fitted to experimental data obtained for a single siRNA targeting the coding region of the EGFP mRNA. The authors obtain the best fit for $\delta(X_m, X_s) = pX_m X_s^h/(q^h + X_s^h)$ with $h \sim 4.5$, $p \sim 0.008/\mathrm{min}$, and $q \sim 0.1\mathrm{pmol}$. Within a similar framework, Osella *et al.* [228] study the so-called miRNA-mediated feedforward loop in which a master transcription factor regulates a miRNA together with a set of target genes,

and the mathematical models studied by Loinger *et al.* [229] additionally account for the concentration level of the argonaute protein complex. Baker *et al.* [230] study analytically the impact of multiple small non-coding RNAs on the regulation of a single target mRNA and subsequently the dynamics of protein production. Ragan *et al.* [106] combine the concept of miRNA binding site accessibility with miRNA and mRNA concentration levels. For $[S]$, $[T]$ and $[ST]$ denoting the equilibrium (final) concentrations of the miRNA, target mRNA and of the hybridized structure, respectively, the authors utilize the equilibrium condition $[ST]/([S][T]) = exp(-\Delta\Delta G/c)$, where $\Delta\Delta G$ is the energy score that accounts for making the binding site accessible and the free energy of the hybridized structure (the constant c stands for the product of the gas constant and the temperature). Combining the equilibrium condition with conservation of mass equations (for initial concentrations $[S_0]$ and $[T_0]$) eventually leads to an analytical expression for $[S]$ in terms of $[S_0]$, $[T_0]$ and $\Delta\Delta G$, where the latter is calculated for a particular binding site. Marin and Vaniček [231] introduce a new accessibility-based algorithm that uses a statistical analysis of all putative binding sites for a given miRNA-3'UTR pair. Among the top 100 target predictions for 153 fruit fly miRNAs, the algorithm finds more than twice as many validated targets compared to other accessibility-based target prediction methods. Reviews of existing miRNA target prediction tools and information about latest developments can be found in [232] and [233]. The target prediction tool `CoMeTa` designed by Gennarino *et al.* [234] operates on the assumption that targets of a given miRNA are co-expressed with each other. The target prediction score is based upon the evaluation of thousands of publicly available microarray data. For the 675 human miRNAs analyzed in the study, more than 90% of the validated targets fall within the first 50% of predicted targets (which, however, could be a large number). In a similar way, the tool `miRror` designed by Balaga *et al.* [235] combines scores produced by an ensemble of established miRNA target prediction tools with rankings obtained from gene expression and HITS-CLIP data.

Johnson and Srivastava [96] consider the problem of selecting an antisense sequence that is able to effectively bind to a target mRNA and block protein synthesis. One of the key features of the authors' method is the presuppositions that mRNA secondary structures are in a constant state of flux and are

assuming different suboptimal states, which also determines the approach taken in the present work. Johnson and Srivastava [96] designed a tool that generates and compares suboptimal secondary structures of a given mRNA sequence. The comparison aims at identifying regions that are least similar among the set of folded structures, which indicates volatility in intramolecular hydrogen bonding. Such regions are seen as candidates for antisense binding. The method is evaluated on six mRNA sequences and compared to results produced by the `Soligo` application of `Sfold` [236].

In the present work, we study miRNA bindings to sets of metastable secondary structures induced by 3'UTRs and their mutated counterparts, where the SNPs are located within the miRNA binding site. Long *et al.* [97] briefly mention the analysis of miRNA bindings to metastable conformations. The authors compare the prediction performance over samples of 1,000 secondary structures without constraints to samples of 100 metastable conformations. They report the sample of 1,000 conformations is more favorable. In our approach, we discriminate between metastable conformations with regard to the depth and average opening energies.

3.2 Approach

As mentioned in the preceding section, the present study assumes the existence of multiple active RNA conformations instead of a unique minimum free energy conformation as the single biologically active state, see [96], [36], and the literature therein. The second basic feature of our approach relates to the presence of multiple copies of each individual mRNA. In more detail, we proceed as follows for a given [mRNA; SNP; miRNA] instance (a flowchart of the approach is provided in Figure 3.1):

(A) For both alleles (3'UTR and SNP), the sets MS(3'UTR, δE) and MS(SNP, δE) of metastable states within an energy offset δE above the MFE conformation are identified by using `RNAsubopt` and `Barriers`; cf. [37; 237; 238]. Alternative tools are, for example, (modified) version of `RNAlocopt` [40] and `Regliss` [44].

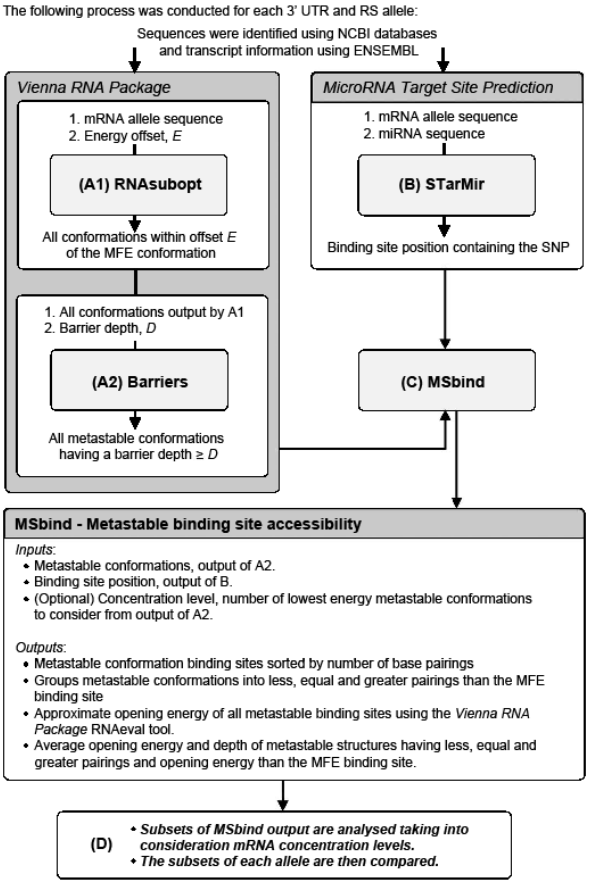


Figure 3.1: Approach of data analysis.

(B) `STarMir` [239] is applied to both alleles for the given miRNA. Although the 14 basic cases [mRNA; SNP; miRNA] selected from recent literature were analyzed in the corresponding publications by prediction tools different from `STarMir`, we obtained for each of the cases at least for one allele a binding site predicted by `STarMir` with the SNP position inside.

(C) For the predicted binding site BS(miRNA), the elements of the sets MS(3'UTR, δE) and MS(SNP, δE) are examined with respect to a number of basic features, such as the number of base pair bindings within BS(miRNA) and the approximate free energy ΔG_{BS} of bindings within BS(miRNA) according to standard data of the Nearest Neighbour Model [26; 240].

(D) Subsets of MS($\cdots, \delta$E) are analyzed in the context of the number of mRNA copies, as analyzed in Subkhankulova *et al.* [104], in the same way as in (C), i.e. for $k = 20, 60, 100$ the sets MS($\cdots, k$) $\subset$ MS($\cdots, \delta$E) are examined, where k indicates the assumption about the number of mRNA copies. We note that the number of copies is different from concentration levels, which are measured, e.g. *per mol*, see [227].

The analysis in this work has been carried out on the first retrieved [mRNA; SNP; miRNA] instances described in detail in Chapter 2- Section 2.2. These fourteen cases are the following:

[LIG3; rs4796030; miR-221].
[CBR1; rs9024; miR-574-5p].
[HTR3E; rs56109847; miR-510-5p].
[HLA-G; rs1063320; miR-148a-3p].
[PARP1; rs8679; miR-145-5p].
[WFS1; rs1046322; miR-668-3p].
[IL-23R; rs10889677; let-7e].
[RYR3; rs1044129; miR-367].
[AGTR1; rs5186; miR-155-5p].
[FGF20; rs12720208; miR-433-3p].
[HOXB5; rs9299; miR-7-5p].
[RAD51; rs7180135; miR-197-3p].
[ORAI1; rs76753792; miR-519a-3p].
[RAP1A; rs6573; miR-196a].

In these 14 cases, the miRNA bindings are predicted by the latest version of `STarMir` [239] at least for the allele with the stronger inhibitory effect with the SNP position being inside the binding region.

3.3 Results

When executing step (A), we applied the standard settings of `RNAsubopt` and `Barriers`, where isolated base pairings are not permitted and free energy values are discriminated with an accuracy of 0.1 kcal/mol. The setting of δE depends

on the length of the 3'UTR and was selected in such a way that a sufficiently large number of metastable conformations is available for analysing $\mathrm{MS}(\cdots, k)$ with $k \leq 100$. Along with the energy offset δE, we tried to restrict metastable states to 'deep' local minima. The parameter D indicates the 'depth' of a local minimum or 'escape height' from a local minimum, which is taken from the barrier tree as the distance to the nearest saddle point. By $|\mathrm{SecStruc}_{w/s}|$ (short for both cases of $|\mathrm{SecStruc}_{weak}|$ and $|\mathrm{SecStruc}_{strong}|$) we denote the number of secondary structures returned by `RNAsubopt` for an offset δE above the MFE conformation, where the index w indicates the allele with the weaker and s with the stronger miRNA inhibitory effect (binding prediction - for [ORAI1;rs76753792;miR-519a-3p]). Analogously, $|\mathrm{MS}_{w/s}|$ is the number of local minima within the δE range with an 'escape height' larger or equal to D.

3.3.1 Number of Metastable Conformations

The results of step (A) with respect to δE,D , $|\mathrm{MS}_{w/s}|$ and $|\mathrm{SecStruc}_{w/s}|$ are summarized in Table 3.1. For 9 of the 14 instances, the values of $|\mathrm{MS}_w|$ and $|\mathrm{MS}_s|$ are relatively close or (much) larger for $|\mathrm{MS}_s|$, see Table 3.1 and Figure 3.2. For the remaining instances, the ratio $|\mathrm{MS}_w|/|\mathrm{MS}_s|$ ranges from 1.31 (FGF20) up until 6.47 (AGTR1). Thus, the number of secondary structures classified as metastable conformations *per se* does not discriminate between the two cases of weaker and stronger binding to the associated miRNA. The correlation between $|\mathrm{MS}_w|$ and $|\mathrm{MS}_s|$ does not necessarily extend to $|\mathrm{SecStruc}_{w/s}|$ (see Figure 3.3). For example, for LIG3 we have $|\mathrm{MS}_w|/|\mathrm{MS}_s| > 1$, whereas $|\mathrm{SecStruc}_w|/|\mathrm{SecStruc}_s| < 1$.

3.3.2 MicroRNA Binding Sites and Energy Predictions

The results obtained in step (B) are summarized in Table 3.2. For a given input [3'UTR/SNP;miRNA], `STarMir` returns a large number of data items and graphical representations of miRNA binding patterns. We focus on the binding regions and four energy values, all generated by the `STarMir` method:

(a) ΔG_{nucl} relates to the assumption that the initial stage of base-pairing (nucleation) requires a gain in free energy that is greater than the energy

cost for the translational and rotational entropy loss when both miRNA and mRNA are fixed in a conformation by intermolecular base-pairing. The value of ΔG_{nucl} is calculated by using a sample of 1,000 structures computed by `Sfold`, where the calculation is restricted to short base-pair blocks within a 4 nt single-stranded segment of a putative binding site. The energy cost for the translational and rotational entropy loss is called initiation energy ΔG_{init}, and the standard setting in the `STarMir` tool is $\Delta G_{\text{init}} = 4.09$ kcal/mol, i.e. $\Delta G_{\text{nucl}} + \Delta G_{\text{initiation}} < 0$ kcal/mol can be seen as a basic requirement for miRNA–mRNA interaction.

(b) $\Delta G_{\text{disrupt}}$ is the energy needed for the disruption of base pairs that are present within a putative binding site in a given mRNA secondary structure. We set $\Delta\Delta G_{\text{dis}} = \Delta G^{w}_{\text{disrupt}}$ - $\Delta G^{s}_{\text{disrupt}}$.

(c) ΔG_{hybrid} is the energy gained by the hybridisation of the miRNA with the particular binding site.

(d) ΔG_{total} is the basic `STarMir` score defined by $\Delta G_{\text{total}} = \Delta G_{\text{hybrid}}$ - $\Delta G_{\text{disrupt}}$. We set $\Delta\Delta G_{\text{tot}} = \Delta G^{s}_{\text{total}}$ - $\Delta G^{w}_{\text{total}}$.

The binding regions and energy values shown in Table 3.2 are determined by the `STarMir` prediction with the strongest seed match among all predictions having the SNP position inside, leading in some cases to weaker ΔG_{total} values. In case of a missing strong seed match the target predictions provided by the `PITA` tool [24] are taken into account.

By $|\text{BP}_w|$ and $|\text{BP}_s|$ we denote the number of base pairs in the corresponding MFE conformations, which are part of the output obtained in step (A), within the miRNA binding region predicted by `STarMir` for an individual allele (w = *weak* and s = *strong* allele with respect to miRNA binding). Detailed information about the `STarMir` output is provided in Supplementary Material in Section .3, which also includes `PITA` predictions and, where available, data provided by `FindTar` [241]. The `PITA` tool returns miRNA binding predictions for the binding sites covering the SNP position as predicted by `STarMir` for nine out of the fourteen instances: HTR3E, HLA-G, WFS1, IL-23R, RYR3, AGTR1, FGF20, RAD51,

RAP1A. `TargetScan`, Release 6.2 from June 2012, returns miRNA binding predictions for the 3'UTR of HTR3E, RYR3, AGTR1, FGF20 and ORAI1, however, with the SNP position located at some distance from the actual binding region predicted by `TargetScan`. For the `miRanda` tool, version v3.3a, we lowered the cut-off score from standard 140 to 90 and obtained for thirteen instances binding predictions for at least one allele with the SNP position inside. For CBR1, `miRanda` returns a binding prediction for the A-allele (s-case) with a score of 138, and no prediction is made for the G-allele (w-case). For the remaining twelve instances, the binding predictions for the four instances PARP1, RYR3, HOXB5 and ORAI1 are inconclusive (identical or very close scores for both alleles). For the eight instances HTR3E, HLA-G, WFS1, IL-23R, AGTR1, FGF20, RAD51 and RAP1A, the miRNA binding prediction is stronger for the allele with the stronger binding claimed in the underlying literature source (minimum score difference equal to 16, maximum 32), i.e., overall, 64% of the experimental data are supported by the `miRanda` tool.

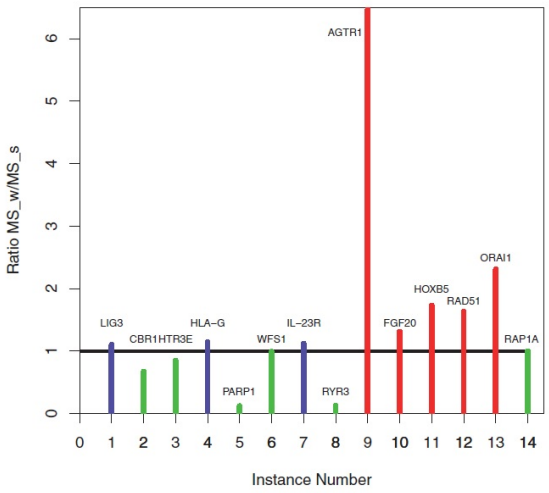


Figure 3.2: $|\mathrm{MS}_{weak}|/|\mathrm{MS}_{strong}|$ ratios. We obtained a larger or equal number of metastable conformations for the allele with the stronger miRNA binding in the case of CBR1, HTR3E, PARP1, WFS1, RYR3 and RAP1A (green colour). The ratio is larger but close to one in the case of LIG3, HLA-G and IL-23R (blue colour). In the case of AGTR1, FGF20, HOXB5, RAD51 and ORAI1, we obtained a much larger number of metastable conformations for the allele with the weaker miRNA binding (red colour).

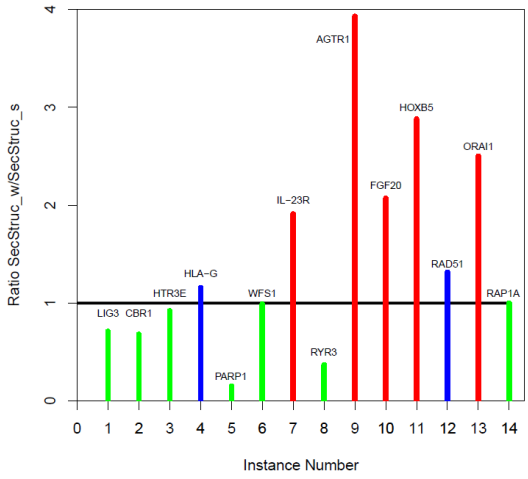


Figure 3.3: $|\text{SecStruc}_{weak}|/|\text{SecStruc}_{strong}|$ ratios. In the case of LIG3, CBR1,HTR3E, PARP1, WFS1, RYR3 and RAP1A, we obtained a larger or equal number of secondary structures calculated by RNAsubopt for the allele with the stronger miRNA binding (green colour). For HLA-G and RAD51 (blue colour), the ratio is larger but close to one, and we obtained a much larger number of secondary structures for the allele with the weaker miRNA binding in the case of IL-23R, AGTR1, FGF20, HOXB5 and ORAI1 (red colour).

Table 3.1: Data returned by `RNAsubopt` and `Barriers`

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3'UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
δE kcal/mol	6.0	6.0	6.0	4.0	3.0	2.7	2.0	2.3	2.3	2.3	2.0	2.2	2.0	2.0
D kcal/mol	1.2	1.4	1.4	1.4	1.2	0.8	1.2	1.2	0.8	0.8	0.8	1.2	0.9	0.9
$\lvert\text{SecStruc}_w\rvert$	2.1×10^4	1.1×10^7	2.4×10^5	9.7×10^6	1.6×10^6	1.1×10^7	7.6×10^5	5.0×10^6	2.1×10^6	2.5×10^6	1.4×10^6	1.2×10^7	5.5×10^5	1.7×10^6
$\lvert\text{SecStruc}_s\rvert$	2.9×10^4	1.6×10^7	2.4×10^5	8.4×10^6	1.1×10^7	1.1×10^7	4.0×10^5	1.4×10^7	5.5×10^7	1.2×10^6	4.8×10^5	9.0×10^6	2.2×10^5	1.7×10^6
$\lvert\text{MS}_w\rvert$	349	7457	996	11957	1709	79273	1080	2628	14943	7783	6481	6291	1332	238
$\lvert\text{MS}_s\rvert$	317	11187	1173	10473	14281	79577	964	19936	2308	5936	3746	3850	577	238

Table 3.2: MicroRNA binding predictions by `STarMir`

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3'UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
miR-	221	574	510	148a	145	668	let-7e	367	155	433	7	197	519a	196a
SNP pos	83	133	76	233	607	253	309	839	86	182	141	718	86	366
BindSite-w	77-98	121-62	50-80	221-39	592-614	234-58	291-308	830-57	57-90	166-87	126-54	707-25	81-102	348-70
BindSite-s	80-98	121-62	50-80	221-39	592-614	234-58	291-310	835-57	57-90	166-87	126-54	707-25	69-88	348-70
$\Delta G^w_{\text{total}}$	-8.53	-19.72	-14.24	-14.62	1.70	-6.40	0.70	4.35	11.28	-2.18	-8.29	-7.00	5.62	-11.30
$\Delta G^w_{\text{disrupt}}$	-11.77	-8.09	-8.26	-9.28	-20.90	-14.20	-21.40	-19.65	-27.89	-10.12	-13.61	-15.00	-23.52	-5.40
$\Delta G^w_{\text{hybrid}}$	-20.30	-27.80	-22.50	-23.90	-19.20	-20.60	-20.70	-15.30	-16.60	-12.30	-21.90	-22.00	-17.90	-16.70
ΔG^w_{nucl}	-3.39	-5.36	-2.46	-4.55	-0.65	-1.72	-0.05	-0.19	-2.27	-4.66	-0.49	-1.20	-0.02	-3.97
$\Delta G^s_{\text{total}}$	-8.64	-21.84	-16.05	-19.49	-7.51	-15.96	1.12	-3.38	7.66	-5.23	-11.20	-14.31	-7.21	-16.25
$\Delta G^s_{\text{disrupt}}$	-15.96	-5.96	-12.65	-11.01	-12.19	-10.64	-25.82	-11.42	-28.56	-9.27	-11.20	-14.29	-9.49	-5.05
$\Delta G^s_{\text{hybrid}}$	-24.60	-27.80	-28.70	-0.50	-19.70	-26.60	-24.70	-14.80	-20.90	-14.50	-22.40	-28.60	-16.70	-21.30
ΔG^s_{nucl}	-0.20	-5.22	-5.05	-3.24	-0.93	-4.42	0.00	-1.22	-0.58	-2.36	-1.39	-0.56	-4.32	-6.97
$\Delta\Delta G_{\text{tot}}$	-0.11	-2.12	-1.81	-4.87	-9.21	-9.56	0.42	-7.73	-3.62	-3.05	-2.91	-7.31	-12.83	-4.95
$\Delta\Delta G_{\text{dis}}$	4.19	-2.13	4.39	-12.89	-8.71	-3.56	4.42	-8.23	0.67	-0.85	-2.41	-0.71	-14.03	-0.35
$\lvert\text{BP}_w\rvert$	9	14	9	9	22	12	15	14	24	10	15	6	16	12
$\lvert\text{BP}_s\rvert$	7	12	15	11	11	12	17	7	22	9	15	6	6	12

Except for IL-23R (L=851), the ΔG_{total} predictions by `STarMir` are stronger for the allele identified for stronger miRNA bindings in the corresponding publication (cf. the row for $\Delta\Delta G_{\text{tot}}$ and description of instances in Chapter 2; we recall that for ORAI1 no particular miRNA is mentioned). Except for LIG3 (L=124), HTR3E (L=302), HLA-G (L=386) , IL-23R (L=851) and AGTR1 (L=888), the absolute value of $\Delta G_{\text{disrupt}}$ is smaller for the allele with the stronger miRNA binding stated in the corresponding publication (s-allele indicated in Table 3.2, see also the row for $\Delta\Delta G_{\text{dis}}$). For IL-23R and AGTR1, the total `STarMir` score is positive for both alleles, for IL-23R even with $\Delta G^{s}_{\text{total}} > \Delta G^{w}_{\text{total}}$.

The instance [IL-23R;rs10889677;let-7e] has been analyzed in [202], with strong experimental evidence for an inhibitory effect of let-7e on the C-allele. The binding patterns for the C allele provided by `STarMir` (see Supplementary Material in Section .3) and in Figure 3A of [202] differ only slightly towards the 5' end of the C allele. We note that for the A allele we have the only case in Table 3.2 where the SNP position is not within the 'binding site' predicted by `STarMir`. The `PITA` tool returns (the equivalent of) $\Delta G_{\text{total}} = -14.12\,\text{kcal/mol}$ for the C allele, and $\Delta G_{\text{total}} = -10.02\,\text{kcal/mol}$ for the A allele (see Supplementary Material in Section .3), which is in line with the experimental data from [202]. The absolute value of $\Delta G_{\text{disrupt}}$ is also slightly smaller for the C allele.

For [LIG3;rs4796030;miR-221], the absolute value of $\Delta G_{\text{disrupt}}$ is larger for the C allele by 4.19 kcal/mol, and there is no `PITA` prediction with the SNP position inside the reported binding region. The `PITA` predictions close to the SNP position favor the C allele with respect to the total score and ΔG_{hybrid}, which complies with [182]. The selection of positions 77–98 for the A allele and 80–98 for the C allele is motivated by the seed-like bindings predicted by `STarMir`, see Supplementary Material in Section .3.

For ORAI1, the positions 69–88 were selected due to $\Delta G^{s}_{\text{nucl}} = -4.32\,\text{kcal/mol}$. For [HTR3E;rs56109847;miR-510-5p], the ΔG_{hybrid} value is stronger for the G allele (s-case) by 6.2 kcal/mol, and the stronger binding prediction is supported by `PITA` and `FindTar` (no binding prediction for A allele-w-case). For [HLA-G;rs1063320;miR-148a-3p], `PITA` returns a stronger total score for the G allele, which is in line with the `STarMir` total score and experimental data from [189]. However, as for `STarMir`, the absolute value of $\Delta G_{\text{disrupt}}$ is larger for the G allele

(by 1.12 kcal/mol for PITA and 1.73 kcal/mol for STarMir). FindTar also strongly supports the miRNA binding to the G allele.

For [AGTR1;rs5186;miR-155-5p], the ΔG_{hybrid} value is stronger for the A allele (s-case) by 4.3 kcal/mol. The stronger binding prediction for the A allele is supported by PITA and FindTar (no binding prediction for C allele-w-case).

For the 14 instances under consideration, the PITA predictions either support (or improve for IL-23R) the STarMir predictions displayed in Table 3.2, or no predictions with the SNP position inside the reported binding region are returned. In summary, if STarMir and PITA (for IL-23R) are taken together, the ΔG_{total} predictions are stronger for the corresponding allele (s-case) identified in the description of the 14 instances in Chapter 2.

3.3.3 Analysis of Metastable Conformations

STarMir and PITA operate on sequences as input, not on representations of secondary structures. Therefore, we utilize RNAeval [237] for the energy evaluation of metastable conformations within binding regions predicted by STarMir, which also complies with the data generated by RNAsubopt and Barriers, see Table 3.1. To facilitate a coherent analysis of energy values, we use energy values calculated for binding regions within the corresponding MFE structure as templates for comparisons, instead of using the associated $\Delta G_{\text{disrupt}}$ values reported in Table 3.2.

Table 3.3: Energy values calculated by `RNAeval`

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3'UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
$\Delta\Delta G_{\text{mfe}}$	-2.20	1.00	-13.51	-0.80	13.02	0.00	-4.20	0.75	2.70	3.20	-0.70	-1.90	21.81	0.00
$\Delta\Delta G_{\text{tot}}$	-1.95	1.31	-11.18	-0.80	17.98	0.03	-7.44	0.82	2.74	2.83	-0.56	-1.85	21.80	0.00
$\Delta\Delta G_{100+}$	-2.60	2.19	-11.26	-1.18	12.92	0.00	-4.20	0.58	2.70	3.91	-0.65	-1.90	21.81	0.00
$\Delta\Delta G_{60+}$	-3.20	2.10	-11.39	-1.73	11.48	0.00	-4.20	0.66	2.70	4.19	-0.65	-1.90	21.81	0.00
$\Delta\Delta G_{20+}$	-3.07	3.56	-10.28	-1.61	7.03	0.00	-4.20	0.75	2.70	4.67	-0.70	-1.90	21.81	0.00
$\Delta\Delta G_{\text{asc:}100+}$	-1.90	1.04	-6.86	-0.69	21.84	7.42	-13.40	2.00	2.72	1.84	1.50	-2.18	21.65	0.00
$\Delta\Delta G_{\text{asc:}60+}$	-1.58	0.98	-4.70	-0.78	21.36	7.48	-13.40	2.13	2.72	1.93	3.38	-2.37	21.54	0.00
$\Delta\Delta G_{\text{asc:}20+}$	-0.91	0.93	-4.38	-0.69	21.04	7.62	-13.40	2.61	2.70	2.00	2.36	-2.61	21.00	0.00
R_{100+}	0.96	0.90	1.14	1.00	0.92	1.00	0.98	0.92	1.25	1.06	1.06	1.00	1.14	1.00
R_{60+}	0.93	0.92	1.11	1.00	0.93	1.00	1.00	0.91	1.32	1.05	1.05	1.00	1.10	1.00
R_{20+}	0.88	0.91	1.20	1.01	0.93	1.00	1.00	0.93	1.09	1.04	1.04	1.00	1.07	1.00
$\Delta\text{D}_{\text{open}}^{100+}$	0.05	-0.05	0.15	0.02	0.35	-0.08	0.30	0.04	1.13	-0.07	0.17	0.23	-0.05	0.00
$\Delta\text{D}_{\text{open}}^{60+}$	0.05	-0.11	0.26	0.49	0.53	-0.10	0.40	0.00	1.13	-0.11	-0.14	0.39	0.08	0.00
$\Delta\text{D}_{\text{open}}^{20+}$	0.04	0.00	0.13	-0.02	0.64	-0.16	0.40	-0.26	1.13	-0.15	-0.03	1.05	0.75	0.00

Let S denote a secondary structure (either MFE conformation or metastable conformation) for a 3'UTR (wild-type or SNP-type) listed in Table 3.1. By S_{open} we denote the associated secondary structure where all base pair bindings within the miRNA binding region reported in Table 3.2 are removed. For example, if S is the MFE conformation of the 3'UTR of the C allele of LIG3, then S has seven base pair bindings in positions 80–98 (see Table 3.2), and the seven base-pair bindings are removed in S_{open}. For S being a metastable conformation, the number of base-pair bindings within the miRNA binding region can be larger, the same or smaller in comparison to the corresponding value reported in Table 3.2 for the MFE conformation. We then define

$$\Delta G^{\text{ind}}_{\text{open}} = \texttt{RNAeval}(S) - \texttt{RNAeval}(S_{\text{open}}); \tag{3.1}$$

$$\Delta\Delta G_{\text{ind}} = \Delta G^{\text{ind:}s}_{\text{open}} - \Delta G^{\text{ind:}w}_{\text{open}}. \tag{3.2}$$

The index 'ind' specifies the different cases we consider, and the different values in (3.1) and (3.2) relate either to individual structures or to average values (according to the value assigned to 'ind') over sets of metastable conformations:

(i) ind = mfe indicates the single MFE conformation.

(ii) ind = tot indicates the average value over all metastable conformations as counted in Table 3.1. For example, $\Delta G^{\text{tot:}s}_{\text{open}}$ stands for the average value over 317 metastable conformations in case of the C allele of LIG3, see Table 3.1.

(iii) ind = N+ indicates for N = 20, 60 and 100 the N+ metastable conformations S with the N+ lowest free energy values calculated by `RNAeval`(S) that are above the MFE conformation.

The output obtained in step (A) consists of a list of metastable secondary structures within the energy offset δE above MFE conformations, ordered with respect to increasing free energy. Each energy level usually adds more than a single conformation. As an example, three conformations are accumulated at energy level -35.0 kcal/mol, in the case of the C allele of LIG3 (see Figure 3.4). Therefore, the notion 'N+' indicates that the highest energy level involved covers N conformations above the MFE structure, plus in most cases some more structures, i.e. the actual number N+ of confor-

mations can be slightly larger than N. For example, for LIG3 C allele, 100+ means 108 metastable conformations as shown in Figure 3.5. Note that for each case, N+ may be different for wild type and variant alleles.

There are a few exceptions for 60+, where the number is between 50 and 60, because the next energy level results already in a conformation number above 100. For example, for PARP1, 60+ means 55, because 55 conformations are accumulated at energy level −186.4 kcal/mol, while level −186.3 kcal/mol adds 48 conformations, which leads to 100+ = 103.

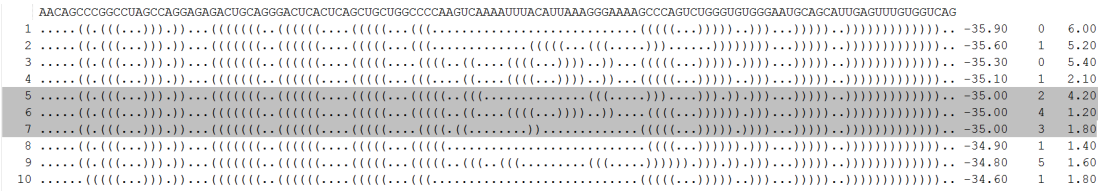


Figure 3.4: Example of the output obtained in step (A) for the case of the LIG3, C allele, where only ten conformations are displayed. The first secondary structure corresponds to the MFE conformation. The three highlighted conformations are accumulated at energy level −35.0 kcal/mol.

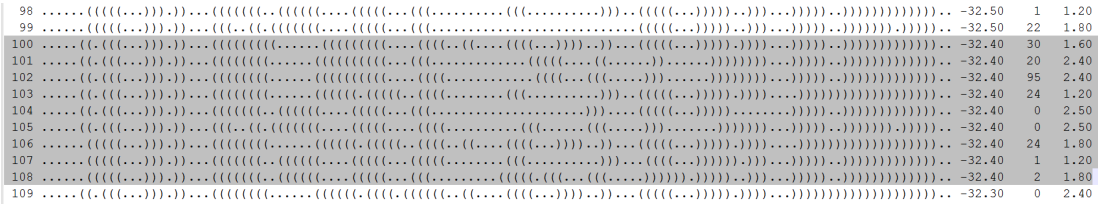


Figure 3.5: Part of the output obtained in step (A) for the case of the LIG3, C allele. Level −32.4 kcal/mol adds the nine highlighted conformations, which leads to 100+ = 108.

Furthermore, we order the metastable conformations S_{ms} with respect to the absolute value of $|\Delta G_{\mathrm{open}}|$ and the depth $D(S_{\mathrm{ms}})$ (deepest first), respectively (see Figure 3.6 for an example). As in (3.2), we define

$$\Delta\Delta G_{\mathrm{asc:N+}} = \Delta G_{\mathrm{open}}^{\mathrm{asc:N+}:s} - \Delta G_{\mathrm{open}}^{\mathrm{asc:N+}:w}; \tag{3.3}$$

$$R_{\mathrm{N+}} = \mathsf{D}_{\mathrm{N+}}^{w}/\mathsf{D}_{\mathrm{N+}}^{s}, \tag{3.4}$$

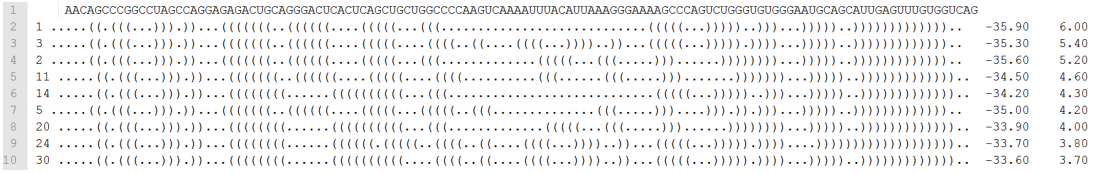


Figure 3.6: Part of the output obtained in step (A) for the case of the LIG3, C allele. The conformations are ordered with respect to the depth $D(S_{\rm ms})$, where the ten deepest metastable conformations are displayed.

where $\mathsf{D}_{\rm N+}^{w/s}$ denotes the average depth of the N+ deepest metastable conformations. The index 'asc' in (3.3) indicates that, unlike in (3.2), the N+ metastable conformations are ranked in ascending order with respect to $|\Delta G_{\rm open}|$, and the average value is taken over N+ conformations.

Finally, we combine opening energies, as defined in (3.2) and (3.3), with the depth $D(S_{\rm ms})$ of metastable conformations, as defined by $\mathsf{D}_{\rm N+}^{w/s}$ and used in (3.4): We look at the average depth of structures $S_{\rm ms}$ with the N+ smallest values of $|\Delta G_{\rm open}|$. Let $\mathsf{D}_{\rm open}^{\rm N+:w/s}$ denote the average depth of metastable conformations $S_{\rm ms}$ counted in the calculation of $\Delta G_{\rm open}^{\rm asc:N+:w/s}$. We then set

$$\Delta\mathsf{D}_{\rm open}^{\rm N+} \;=\; \mathsf{D}_{\rm open}^{\rm N+:s} - \mathsf{D}_{\rm open}^{\rm N+:w}. \tag{3.5}$$

Comprehensive information about the distribution of metastable conformations and their respective energy values is provided in Supplementary Material in Section .4. In Table 3.3 we report representative data that are useful for discriminating between the two alleles involved for each instance. Positive values of $\Delta\Delta G_{\rm mfe}$, $\Delta\Delta G_{\rm tot}$, $\Delta\Delta G_{\rm N+}$, $\Delta\Delta G_{\rm asc:N+}$ and $\Delta\mathsf{D}_{\rm open}^{\rm N+}$ are interpreted as being in favor of the allele with the stronger miRNA binding stated in the underlying literature source, which is also the case for $R_{\rm N+} < 1$. For LIG3 (L=124), HTR3E (L=302), HLA-G (L=386), IL-23R (L=851) and RAD51 (L=978), the data for $\Delta\Delta G_{\rm mfe}$, $\Delta\Delta G_{\rm tot}$, $\Delta\Delta G_{\rm N+}$ and $\Delta\Delta G_{\rm asc:N+}$ shown in Table 3.3 are not in favor of the allele with the stronger miRNA binding stated in the underlying literature (HOXB5 with L=952 not included here, please see later in the text). For LIG3, IL-23R and RAD51 the values of $R_{\rm N+}$ are either close to or smaller than 1.00, and the values

of $\Delta\mathsf{D}^{\mathrm{N+}}_{\mathrm{open}}$ are all positive, suggesting more stable local minima for structures with the smallest absolute value of opening energies for the allele with the stronger miRNA binding.

For HLA-G (L=386), the values of $R_{\mathrm{N+}}$ are equal or close to 1.00, and the values of $\Delta\mathsf{D}^{\mathrm{60+}}_{\mathrm{open}}$ and $\Delta\mathsf{D}^{\mathrm{100+}}_{\mathrm{open}}$ are positive. Moreover, the $\Delta\Delta G_{\mathrm{asc:N+}}$-values shown in Table 3.3 are close to zero, and the values of $\Delta\Delta G_{\mathrm{N+}}$ are in the range of $\Delta G^{s}_{\mathrm{disrupt}} - \Delta G^{w}_{\mathrm{disrupt}} = -1.73\,\mathrm{kcal/mol}$ from Table 3.2. Thus, for HLA-G the values of $\Delta G^{s/w}_{\mathrm{hybrid}}$ shown in Table 3.2 seem to be decisive for an assessment of a putative miR-148a-3p↔HLA-G/rs1063320 binding (based on prediction tools).

For HTR3E, only the values of $\Delta\mathsf{D}^{\mathrm{N+}}_{\mathrm{open}}$ are in favor of the allele with the stronger miRNA binding stated in the underlying literature. Unlike the case of HLA-G, the negative values of$\Delta\Delta G_{\mathrm{N+}}$ and $\Delta\Delta G_{\mathrm{asc:N+}}$ are more substantial, i.e. in terms of absolute values above the range of corresponding values from Table 3.2. Thus, only the values of $\Delta G^{s}_{\mathrm{hybrid}}$, $\Delta G^{s}_{\mathrm{nucl}}$ (both Table 3.2) and $\Delta\mathsf{D}^{\mathrm{N+}}_{\mathrm{open}}$ (Table 3.3) support the binding to the G allele (s-case).

For HOXB5, the $\Delta\Delta G_{\mathrm{N+}}$ values are negative yet close to zero, and the $\Delta\Delta G_{\mathrm{asc:N+}}$ values support the stronger miRNA binding of the A-allele (s-case), which makes the HOXB5 instance different from the five instances discussed earlier in the text. The $R_{\mathrm{N+}}$-values are close to 1.00, and $\Delta\mathsf{D}^{\mathrm{100+}}_{\mathrm{open}}$ is positive.

For RAP1A (L=1078), the `STarMir` predictions for $\Delta G_{\mathrm{disrupt}}$ and $\Delta G_{\mathrm{hybrid}}$, respectively, are very close for both alleles. The SNP at position 366 is located in the middle of a loop (positions 363–369) in both MFE secondary structures, which leads to identical values of $\Delta G^{\mathrm{ind:}s/w}_{\mathrm{open}}$ and related features. Therefore, similar to HTR3E, $\Delta G^{s}_{\mathrm{hybrid}}$ and $\Delta G^{s}_{\mathrm{nucl}}$ from Table 3.2 appear to determine the evaluation of miR-196a↔RAP1A/rs6573 bindings. For CBR1 (L=284), WFS1 (L=779), RYR3 (L=880), AGTR1 (L=888), FGF20 (L=903) and ORAI1 (L=1034), the data for $\Delta\Delta G_{\mathrm{mfe}}$, $\Delta\Delta G_{\mathrm{tot}}$, $\Delta\Delta G_{\mathrm{N+}}$ and $\Delta\Delta G_{\mathrm{asc:N+}}$ (Table 3.3) are all in favor of the allele with the stronger miRNA binding stated in the underlying literature.However, for each of the five instances at least one of the two parameters $R_{\mathrm{N+}}$ and $\Delta\mathsf{D}^{\mathrm{N+}}_{\mathrm{open}}$ does not fully support the predicted binding.

For RYR3 (L=880), only $\Delta\mathsf{D}^{\mathrm{20+}}_{\mathrm{open}} = -0.26\,\mathrm{kcal/mol}$ is clearly not in favor of the predicted binding. Similarly, the values of $\Delta\mathsf{D}^{\mathrm{N+}}_{\mathrm{open}}$ do not support the predicted stronger binding for CBR1. For ORAI1, this is the case for the $R_{\mathrm{N+}}$ values

and $\Delta\mathsf{D}_{\text{open}}^{100+} = -0.05\,\text{kcal/mol}$, but with a relatively strong value of $\Delta\mathsf{D}_{\text{open}}^{20+} = 0.75\,\text{kcal/mol}$. For AGTR1, all three values of $\Delta\mathsf{D}_{\text{open}}^{\text{N+}}$ are clearly in favor of the predicted binding pattern. The instances WFS1 and FGF20 are the only two cases where for all N+ values both parameters do not support the predicted stronger binding. Finally, for PARP1 (L=769) all energy values shown in Table 3.3 support the stronger miRNA binding to the T-allele (s-case). We recall that for the instance ORAI1 from Chang *et al.* [208] no individual miRNA is identified in the literature source. For ORAI1 and miRNA-519a-3p, the two binding sites predicted by `STarMir` intersect only by 8 nt, with a positive value $\Delta G_{\text{total}} = 5.62\,\text{kcal/mol}$ for the T-allele (w-case), which suggests that no binding occurs. Although $R_{\text{N+}} > 1.00$ for all cases of N+ considered, we obtain strong positive values for $\Delta\Delta G_{\text{mfe}}$, $\Delta\Delta G_{\text{tot}}$, $\Delta\Delta G_{\text{N+}}$ and $\Delta\Delta G_{\text{asc:N+}}$, along with the `STarMir` predictions $\Delta G_{\text{total}} = -7.21\,\text{kcal/mol}$ and $\Delta G_{\text{nucl}} = -4.32\,\text{kcal/mol}$ for the C-allele (s-case). Moreover, the data for $\Delta\mathsf{D}_{\text{open}}^{\text{N+}}$ are $\Delta\mathsf{D}_{\text{open}}^{10+} = 0.89\,\text{kcal/mol}$, $\Delta\mathsf{D}_{\text{open}}^{20+} = 0.75\,\text{kcal/mol}$ and $\Delta\mathsf{D}_{\text{open}}^{60+} = 0.08\,\text{kcal/mol}$, which support a binding of miR-519a to the 3'UTR of ORAI1.

3.4 Conclusion

Out of the 14 instances we analyzed, 13 instances are sensitive to the parameters $\Delta\Delta G_{\text{ind}}$, $\Delta\Delta G_{\text{asc:N+}}$, $R_{\text{N+}}$ and $\Delta\mathsf{D}_{\text{open}}^{\text{N+}}$ we introduced in Equation(3.1) until Equation (3.5). For RAP1A (L=1078), slightly larger values of δE did not create differences between basic parameters for both alleles and eventually led to an unmanageable size of data for standard desktop computer configurations.

The absence of experimental data about copy numbers of mRNA transcripts considered in the present study prevents the selection of a particular value of $\Delta\mathsf{D}_{\text{open}}^{\text{N+}}$ (or of the other two highlighted parameters), which is why we considered four representative values of N+ simultaneously, without calculating p-values. The upper bound of N=100+ is motivated by the work provided in [104] where the authors conclude from observed data for about 13,000 genes that the typical number of gene copies lies between 5 and 20, with 85% of genes having less than 100 copies in a single cell.

Table 3.4: Summary of discrimination between WT and SNP alleles

	LIG3	CBR1	HTR3E	HLA-G	PARP1	WFS1	IL-23R	RYR3	AGTR1	FGF20	HOXB5	RAD51	ORAI1	RAP1A
L(3'UTR) nt	124	284	302	386	769	779	851	880	888	903	952	978	1034	1078
W-allele	A	G	A	C	C	A	A	G	C	T	G	A	T	C
S-allele	C	A	G	G	T	G	C	A	A	C	A	G	C	A
`STarMir` (total)	+	+	+	+	+	+	-	+	±	+	+	+	+	+
N+ selection	20+	20+	20+	60+	20+	20+	100+	100+	20+	20+	100+	20+	20+	20+
$\Delta\Delta G_{\text{asc:N+}}$	±	+	-	±	+	+	-	+	+	+	+	-	+	±
$R_{\text{N+}}$	+	+	-	±	+	±	+	+	-	-	-	±	-	±
$\Delta \text{D}^{\text{N+}}_{\text{open}}$	+	±	+	+	+	-	+	+	+	-	+	+	+	±

The data provided in Table 3.3 indicate that $\Delta\Delta G_{\text{mfe}}$, $\Delta\Delta G_{\text{tot}}$, $\Delta\Delta G_{\text{N+}}$ does not necessarily contribute to a deeper insight into miRNA binding patterns to different alleles, see also the discussion of instances in Section 3.3.3. In particular, $\Delta\Delta G_{\text{tot}}$ and $\Delta\Delta G_{\text{N+}}$ are related only to free energy values of metastable structures, which is why a further discrimination by the depth of metastable conformations and the opening energy of binding regions was introduced. Consequently, we focus in the summary of findings presented in Table 3.4 on the values calculated for $\Delta\Delta G_{\text{asc:N+}}$, $R_{\text{N+}}$ and $\Delta\mathsf{D}^{\text{N+}}_{\text{open}}$.

In Table 3.4, the row '`STarMir`(total)' indicates by '+' that the ΔG_{total} score (see Table 3.2) supports the allele with the stronger miRNA binding stated in the underlying literature; $\pm$ indicates $0 < |\Delta G^{s}_{\text{total}}| < |\Delta G^{w}_{\text{total}}|$. As aforementioned, data about estimations of copy numbers are not available for the mRNA transcripts we consider in the present study. To avoid the inclusion of irrelevant data (by averaging or thresholding), we consider a 'best case scenario' for each instance: We select a value of N+ in such a way that the support of the miRNA binding to the allele identified in the underlying literature source (s-case) is maximized. In case of multiple N+ values (for 11 instances same pattern as in Table 3.4 for at least two N+), the smallest N+ is selected and named in Table 3.4 in the row 'N+ selection'.

For the selected N+, the Table 3.4 entry is labeled as positive '+', if $\Delta\Delta G_{\text{asc:N+}}$ and $\Delta\mathsf{D}^{\text{N+}}_{\text{open}}$ are positive, respectively, or $R_{\text{N+}} < 1.00$. If the data are inconclusive (equal or close to 0.00 or 1.00), we use $\pm$. For example, for HLA-G we select N+=60+ and obtain from Table 3.3 the entries for Table 3.4 as follows: $\Delta\Delta G_{\text{asc:N+}} = \pm$, $R_{\text{N+}} = \pm$, and $\Delta\mathsf{D}^{\text{N+}}_{\text{open}} = +$.

Table 3.4 demonstrates that the combined measure $\Delta\mathsf{D}^{\text{N+}}_{\text{open}}$ defined in (3.5) is the best match to the binding predictions, with two inconclusive and two negative values. The inconclusive value of $\Delta\mathsf{D}^{\text{20+}}_{\text{open}}$ for CBR1 is accompanied by two positive values of the other two parameters, and RAP1A is a special instance due to the SNP location, as discussed in Section 3.3.3. The negative values of $\Delta\mathsf{D}^{\text{20+}}_{\text{open}}$ for WFS1 and FGF20 are accompanied by relatively strong positive values of $\Delta\Delta G_{\text{asc:20+}}$. Future research will focus on how to establish a suitable total score that takes into account weighted partial scores provided by $\Delta\Delta G_{\text{asc:N+}}$, $R_{\text{N+}}$ and $\Delta\mathsf{D}^{\text{N+}}_{\text{open}}$. For example, the simple weight selection $-1 \equiv -$, $0 \equiv \pm$ and $1 \equiv +$

applied to Table 3.4 results in a total score of -1 for HTR3E and FGF20, and a total score of 0 for WFS1, RAD51 and RAP1A. However, WFS1 has a strong $\Delta\Delta G_{\text{asc:20+}}$ score, RAD51 has a strong $\Delta \mathsf{D}^{\text{N+}}_{\text{open}}$ score and RAP1A is a special case due to the location of the SNP. The relatively strong score of $\Delta\Delta G_{\text{asc:20+}}$ could outweigh the other two scores. HTR3E remains an ambiguous case, see also the discussion of the HTR3E-miR-510-5p instance in [174].

We hypothesize that an in-depth analysis of metastable conformations based upon parameters such as $\Delta\Delta G_{\text{asc:N+}}$, $R_{\text{N+}}$ and $\Delta \mathsf{D}^{\text{N+}}_{\text{open}}$ can provide useful information for the assessment of putative miRNA-mRNA bindings in the context of SNPs. In the literature sources we researched for the current study, the number of genes and miRNAs exposed to experimental analysis is relatively small, yet each analysis is time-consuming and costly. Examining features of metastable conformations in a preprocessing step of wet lab experiments may improve the confidence about expected miRNA-mRNA bindings. We emphasize that for ORAI1 no specific microRNA is identified by Chang *et al.* [208]. The data we presented support the binding of miR-519a-3p to the 3'UTR of ORAI1 in the region of position 86.

In the present study, the energy values associated with the accessibility of binding sites are approximated by using `RNAeval`, and the range of the binding sites is directly adapted from the miRNA target prediction tool `STarMir`.

Subsequently, we aim at an automated, individual miRNA-mRNA binding prediction algorithm, which continues our work presented in this chapter. We plan to incorporate the approach into a procedure designed for the analysis of miRNA bindings to 3'UTR conformations, using samples of metastable conformations taken from the sets of secondary structures defining the values of $\Delta\Delta G_{\text{asc:N+}}$, $R_{\text{N+}}$ and $\Delta \mathsf{D}^{\text{N+}}_{\text{open}}$.

Chapter 4

MicroRNA Target Prediction Based upon Metastable RNA Secondary Structures

In Chapter 3, we studied the problem of miRNA bindings to metastable secondary structures in the context of SNPs and mRNA concentration levels. Our analysis showed that the number of metastable structures and features of miRNA bindings to metastable conformations could provide additional information supporting the differences in expression levels of mRNAs and their corresponding SNP variants. As a consequence, we concluded that miRNA target predictions using metastable conformations in a pre-processing step of wet lab experiments may improve the confidence about expected miRNA-mRNA bindings.

In this chapter, we present `RNAStrucTar`, a new miRNA target prediction tool that analyses putative mRNA binding sites within 3'UTR secondary structures representing metastable conformations. The first stage consists of generating conformations that can be classified as deep local minima. The second stage incorporates duplex structure prediction through sequence alignment and energy computation. Target site accessibility related to different sets of metastable conformations is also taken into account. An overall interaction score computed from multiple binding sites is returned. The approach is discussed in the context of single nucleotide polymorphisms (SNPs). In Chapter 2, we selected 34 instances

of type [mRNA;SNP;miRNA] reported in recent literature where methods such as PCR and/or luciferase reporter assays are utilised. If the two main scores returned by `RNAStrucTar` are combined, 25 instances are correctly classified according to experimental findings from the literature, with two false classifications and seven indifferent outcomes. We finally prove that predictions improve when additionally combined with `STarMir` and `PITA` results. We note that a part of this work has been published in [242].

4.1 Background

Computational methods perform the predictions based upon features extracted from experimental data. Early target prediction algorithms primarily focus on sequential features, such as `PicTar` [91], `TargetScan` [85], and `miRanda` [115]. Most methods require the target sequence to have a near perfect complementarity to a region in the miRNA sequence, which is named seed and defined by the first 2–8 nucleotides, starting at the 5' end. The prediction accuracy can be improved by taking into account evolutionary conservation of binding sites in both sequences [93]. Thermodynamic stability of miRNA–mRNA duplex structures is also one of the most frequently used features, where the overall change in Gibbs free energy is employed as an indicator of how strongly bound the sequence pair is. More detailed reviews of features used in existing miRNA target prediction tools can be found in [88; 232; 233; 243; 244].

Advanced methods take into account the secondary structure of mRNAs. One of the key concepts underlying such tools is the accessibility of potential binding sites. In reality, there is an energy cost to be considered for freeing base pairing interactions within the mRNA. This feature was first used in the work of Zhao *et al.* [245], where they provide evidence that RNA accessibility is a principal feature of miRNA target recognition. However, they only integrate the MFE structure, generated by `mfold` [12], into their predictions. Subsequent works consider the complete Boltzmann ensemble of all possible structures rather than relying on a single predicted MFE structure. A statistical sample of the ensemble allows sampling estimates of the structural accessibilities around target sites. Our work in Chapter 3 assumes the existence of multiple active RNA conforma-

tions different instead of a unique MFE structure as the single biologically active state. The second basic feature of our work relates to the presence of multiple copies of each individual mRNA. We studied the problem of miRNA bindings to metastable secondary structures in the context of SNPs and mRNA concentration levels. Our analysis showed that the number of metastable structures and features of miRNA bindings to metastable conformations could provide additional information supporting the differences in expression levels of mRNAs and their corresponding SNP variants. As a consequence, miRNA target predictions using metastable conformations in a pre-processing step of wet lab experiments may improve the confidence about expected miRNA-mRNA bindings. `RNAStrucTar`, the method described in the present work, is a new miRNA target prediction tool that analyses putative mRNA binding sites - in contrast to Chapter 3 - by a specific energy evaluation of duplex structures based upon secondary structures representing metastable conformations.

4.2 Approach

There are two main stages in `RNAStrucTar`. The first stage is the generation of metastable conformations, and the second stage comprises of miRNA target prediction based upon an energy assessment that incorporates target accessibility related to an input set of secondary structures.

While the first stage is similar to our analysis discussed in Chapter 3, the remaining steps are different. In Chapter 3, we looked manually at the range of one single binding site for each case, predicted by `STarMir` with the SNP positions inside. Moreover, the energy values associated with the accessibility of the binding sites were approximated manually by using `RNAeval`. Conversely, `RNAStrucTar` is a self contained tool which analyses automatically the input mRNA sequence for putative binding sites, uses structural accessibility feature related to different sets of metastable conformations and finally takes the number of target sites into account to generate a final score. A flowchart showing the particular steps of `RNAStrucTar` is given in Figure 4.1.

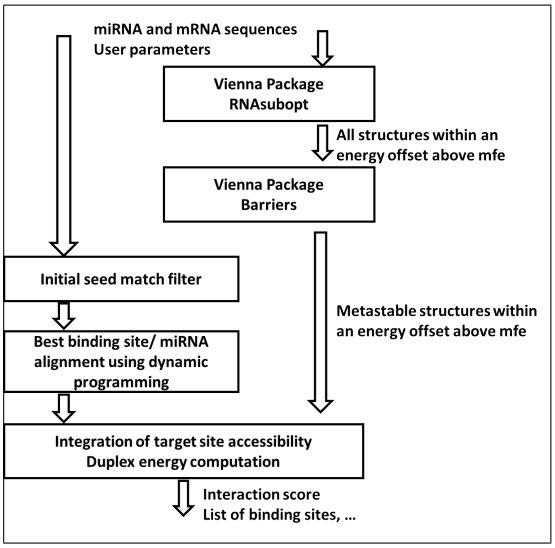


Figure 4.1: RNAStrucTar flowchart.

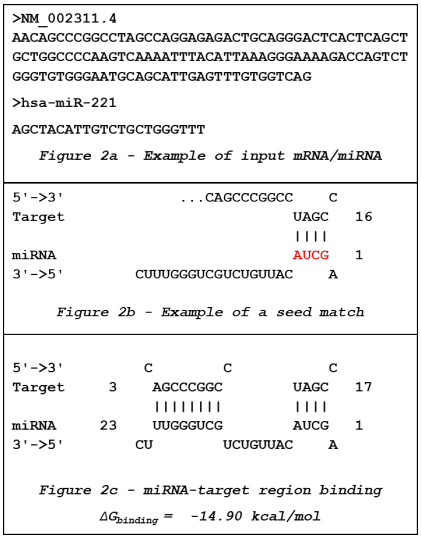


Figure 4.2: RNAStrucTar example.

4.2.1 Metastable Secondary Structures

Metastable secondary structures are generated by using standard tools provided by the Vienna RNA server [237]. The `RNAsubopt` tool by Wuchty *et al.* [37] generates all suboptimal foldings of a sequence in a partial energy landscape defined by an energy range δE above the MFE structure. A method to elucidate the basin structure of landscapes by means of tree-structures representing local minima and their connecting saddle points is provided by the `Barriers` program [39]. The input to `Barriers` is a list of conformations sorted by energy values. The `RNAsubopt` tool together with the `Barriers` program allows the user to identify the set of metastable conformations MS within an energy range δE above the MFE conformation. We denote the number of local minima by ms = |MS|.

4.2.2 Identification of Putative Nucleation Sites

`RNAStrucTar` first scans the mRNA sequence in search for putative nucleation sites, considering complementarity with the seed region of the miRNA. This seed match step is commonly used and considered as a speed-up factor that accelerates the algorithm while it differs from tool to tool in terms of number of of matches.

A flexible miRNA seed window - nucleotides 2 to 8, counting from the 5' end of the miRNA - is used to scan the mRNA sequence for potential target sites. We first predefine the different complementary sequences of the seed region of the miRNA. The number of these sequences depends on the required user defined seed match n-mer (n varies between 3 for shorter seed matches and 7 for longer ones). There must be n consecutive Waston Crick basepairs with no G-U wobble allowed to appear. Each complementary sequence will slide through the input UTR sequence to find all matching positions. All results shown in this work were obtained by using 3-mers or 4-mers complementarity to miRNA positions 2–5 in the seed match step; see Figure 4.2.

4.2.3 [binding region, miRNA]-duplex Structure Prediction

After the seed regions are identified, the upstream flanking region of the seed region is extracted for the next step. Among common features of prediction programs are dynamic programming and the alignment of the miRNA seed region to the target mRNA. We propose a dynamic programming approach for finding minimum energy alignments between the full length of the miRNA and the target sequence for each putative binding site. For this purpose, a modified version of `RNAduplex` [125] is adapted to compute the optimum duplex structure for each putative binding site. For each such site j, the binding pattern and its free energy $\Delta G^j_{\text{binding}}$ are computed according to the seed alignment from the previous step; see Figure 4.2. At the end of this step, weak binding sites are filtered out by applying an energy threshold ϑ with the default setting -10kcal/mol. We considered the performance of various threshold values in filtering out the weak sites that will not go through the next step and we found a cut-off value between -8kcal/mol and -10kcal/mol most suitable to distinguish between confident binding sites and weaker ones. However, users may wish to lower the number of putative binding sites at the end of this step by choosing a more stringent cut-off threshold. This way we obtain k binding sites that satisfy the condition $\Delta G^j_{\text{binding}} \leq \vartheta$, $j = 1, ..., k$.

4.2.4 Integration of Target Site Accessibility

Similar to [24] and [97], we adopted the simplifying assumption that the binding of a miRNA to a longer target mRNA should cause a local structural alteration at the target site, but has no long-range effects on the overall target secondary structure. This leads to a breakage of intramolecular bonds within the target region. For each input secondary structure $\mathrm{F}_i \in \mathrm{MS}$, $i = 1, ..., \mathrm{ms}$, the energy contribution $\Delta G^j_{\mathrm{open},i}$ of the deleted bindings is computed for each site j by using `RNAeval` [237] and according to standard data of the Nearest Neighbour Model [26; 240]. Given F_i, we denote by $\mathrm{F}_{\mathrm{open},i}$ the associated secondary structure where all base pair bindings within j are removed. We then define

$$\Delta G^j_{\mathrm{open},i} \quad = \quad \texttt{RNAeval}(\mathrm{F}_i) - \texttt{RNAeval}(\mathrm{F}_{\mathrm{open},i}). \tag{4.1}$$

4.2.5 miRNA-target Score Derived from a Single Binding Site

At this stage, we estimate the free energy of the miRNA:mRNA duplex structure by using `RNAeval`. For each putative binding site j and for each input secondary structure F_i, we generate an artificial RNA sequence that consists of the original mRNA (3'UTR) sequence, a linker sequence XXXX, and the miRNA sequence. The corresponding folding is denoted by $\mathrm{F}^j_{\mathrm{concat},i}$. The score $\mathrm{S}(\mathrm{miRNA,3'UTR,F}_i,j) = S_{i,j}$ is then defined by

$$S_{i,j} \quad = \quad \texttt{RNAeval}(\mathrm{F}^j_{\mathrm{concat},i}) - \texttt{RNAeval}(\mathrm{F}_i). \tag{4.2}$$

We integrate the scores of multiple conformations F_i for the $h_j \leq \mathrm{ms}$ negative values of $S_{i,j} < 0$ by setting a logarithmic sum of Boltzmann-weighted energies for these different conformations

$$S_j \quad = \quad -\log \sum_{s=1}^{h_j} e^{-S_{i_s,j}}, \; j = 1, ..., k; \tag{4.3}$$

see Figure 4.3. In Figure 4.3 we assume for simplicity $h_j = \mathrm{ms}$ for all $j \leq k$.

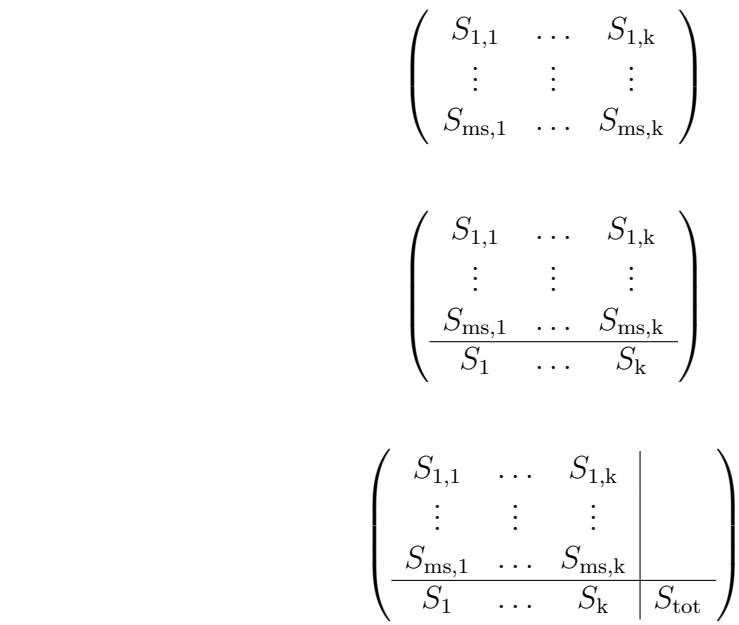


Figure 4.3: Integration of multiple binding sites and conformation into a single score.

The setting according to Eqn. 4.3 is inspired by the `PITA` score [24], where the sum corresponds to the partition function of the conformations, and the logarithmic is therefore linearly related to the probability of that ensemble of conformations.

4.2.6 MicroRNA Target Prediction Scores

Some existing target prediction methods check the presence of multiple target sites and take the number of target sites into account for a final score.

We explored different ways to account for the occurrence of multiple binding sites and we ended up with a scoring function where the emphasis is on combining strong duplex conformations with a user-defined target region. To integrate multiple sites with S_j-scores for a given miRNA and a fixed 3'UTR into an overall

miRNA:target interaction score, we define

$$S_{\text{tot}} \;=\; -\log \sum_{j=1}^{k} e^{-S_j}. \tag{4.4}$$

We note that by definition $S_j < 0$ (assuming $h_j \geq 1$) for all $j \leq k$, see Eqn. 4.3. We recall that each $S_j \leq 0$ represents information about $h_j \leq$ ms bindings to metastable conformations F_i, $i = 1, ..., h_j$, which justifies the notation S_{tot} as total score.

Additionally, other alternative scoring functions were also analyzed: For each conformation F_i, the linear sum S_i of $k_i \leq k$ values of $S_{i,j} < 0$ is computed, and for $h \leq$ ms values of $S_i < 0$, the average value is denoted by S_{sum}. Thus, we define

$$S_i \;=\; \sum_{t=1}^{k_i} \mathrm{S}_{i,j_t}; \tag{4.5}$$

$$S_{\text{sum}} \;=\; \frac{\sum_{s=1}^{h} S_{i_s}}{h}. \tag{4.6}$$

In addition to S_{tot} and S_{sum}, `RNAStrucTar` allows the user to calculate a score S_{u} derived from a binding site u that contains a user defined position (usually, where the SNP is located) within the input RNA sequence.

$$S_{\text{u}} \;=\; \frac{\sum_{s=1}^{h_u} S_{i_s,u}}{h_u}. \tag{4.7}$$

Again, similar to the `PITA` score [24], we define

$$S_{\text{P}} \;=\; \frac{\sum_{s=1}^{h} -\log \sum_{t=1}^{k_i} e^{-S_{i_s,j_t}}}{h}. \tag{4.8}$$

We emphasise that based upon Eqns. 4.3–4.8 the values of S_j, S_{tot}, S_i, S_{sum}, S_{u}, and S_{P} are either negative or not defined (e.g., if $h_j = 0$ for some j or $k_i = 0$ for some i).

4.2.7 Metastable Conformations Sets

In our analysis in Chapter 3, along with the energy offset δE above the MFE conformation, we tried to restrict metastable states to deep local minima. The parameter D indicates the depth of a local minimum or - in other terms - the escape height from a local minimum, which is taken in barrier trees as the distance to the nearest saddle point. We found that out of the three different parameters we introduced, the average depth and the average opening energy of metastable conformations may provide supporting information for a stronger separation between miRNA bindings to the two alleles defined by a given SNP. Here, we aim at individual miRNA–mRNA binding predictions over samples of metastable conformations defined by these parameters. Therefore, we order the metastable conformations with respect to:

(a) The depth D(F_i) in descending order (deepest first).

(b) The absolute value of the opening energy $\Delta G^u_{\text{open},i}$ of the user defined target region, ranked in ascending order.

We obtain the following two sets, where N is a user defined parameter:

(a) **Set A**: the N deepest metastable conformations among MS.

(b) **Set B**: the N most easily accessible conformations in the user defined target region among the deepest metastable conformations.

4.3 Results

4.3.1 Test Dataset

The tools `RNAsubopt` and `Barriers` generate a huge amount of secondary structures, even for a small offset δE above the MFE value. The number of structures returned by `RNAsubopt` grows exponentially with both sequence length and energy range [42]. Consequently, a large scale test or a genome wide prediction analysis is not possible at this stage. In order to test our approach, we use the same data acquisition method as in Chapter 3, i.e., we use miRNA–mRNA pairs

from published experimental work where SNPs are linked to specific diseases. SNPs can be located in miRNA binding regions, and consequently they could affect gene expression. `RNAStrucTar` can be used to evaluate how SNPs affect miRNA regulation by using as input the wild type and the SNP variant. Our aim is to determine the ability of `RNAStrucTar` to provide supportive information for a stronger discrimination between miRNA bindings to the two alleles defined by a given SNP (also denoted as wild type and SNP sequences). We analyzed the 34 instances of [mRNA/3'UTR;SNP;miRNA] interactions described in Chapter 2-Section 2.2.

All results shown in this work were obtained using 3-mers or 4-mers complementarity to the miRNA positions 2–5 in the seed match step. The setting of δE depends on the length of the 3'UTR and was selected in such a way that a sufficiently large number of metastable conformations is available. Experimental findings suggest that the typical number of gene copies lies between 5–20. Therefore, tests were carried out with N=10 and N=20. For each case, the SNP position was used as the user defined position in order to obtain the score S_u.

4.3.2 Energy Scores

We note that the publications of experimental work where the test data are taken from differentiate for each allele pair between *weaker* bindings (expression levels) and *stronger* bindings for the miRNA under consideration. Consequently, we calculate energy scores for the weaker and stronger allele, respectively, where it depends on the particular instance which one of the wild type or SNP sequence produces the stronger or weaker interaction. Thus, for a given input [mRNA/3'UTR;SNP;miRNA], `RNAStrucTar` returns the scores S_{tot} from Eqn. 4.4, S_{sum} from Eqn. 4.6, S_{P} from Eqn. 4.8, and S_{u} from Eqn. 4.7 (binding site u contains SNP position) for two alleles, and subsequently the differences are calculated:

$$\Delta S_{\text{tot}} = S_{\text{tot}}^{\text{stronger}} - S_{\text{tot}}^{\text{weaker}}; \tag{4.9}$$

$$\Delta S_{\text{sum}} = S_{\text{sum}}^{\text{stronger}} - S_{\text{sum}}^{\text{weaker}}; \tag{4.10}$$

$$\Delta S_{\text{P}} = S_{\text{P}}^{\text{stronger}} - S_{\text{P}}^{\text{weaker}}; \tag{4.11}$$

$$\Delta S_{\mathrm{u}} \quad = \quad S_{\mathrm{u}}^{\mathrm{stronger}} - S_{\mathrm{u}}^{\mathrm{weaker}}. \tag{4.12}$$

Negative values of ΔS are expected for a target prediction to be classified as correct. Here, we focus on Case A (applied to Set A as described in Section 4.2.7), although Case B (applied to Set B as described in Section 4.2.7) is briefly discussed. The results obtained for the 34 instances and Case A by using 3-mers complementarity regarding the seed match and with setting N = 10 are summarized in Table 4.1.

Table 4.1: Summary of miRNA binding prediction by `RNAStrucTar`. A'+' ('-') indicates that the score supports the allele with the stronger (weaker) miRNA binding, with ΔS threshold -1kcal/mol for '+'. A '0' means $-1\mathrm{kcal/mol} < \Delta S \leq 0\mathrm{kcal/mol}$.

	L(3'UTR) nt	W-allele	S-allele	miRNA	SNP pos	ΔS_{sum}	ΔS_{u}	`PITA`	`STarMir`
LIG3	124	A(w)	C(v)	221	83	+	+	+	+
MSLN	132	A(v)	G(w)	611	69	+	+	0	-
CGA	258	T(w)	A(v)	1302	215	+	+	+	+
CBR1	284	G(w)	A(v)	574	133	+	+	+	+
HTR3E	302	A(v)	G(w)	510	76	-	-	+	+
KRT81	342	C(v)	G(w)	17	102	+	+	+	-
SPI1	369	T(v)	C(w)	569	330	0	0	+	+
HLA-G	386	C(w)	G(v)	148a	233	+	+	+	+
MTHFD1L	393	A(w)	G(v)	197	120	+	+	+	-
NFKBIA	502	G(w)	A(v)	449a	126	+	0	0	+
TYMS	502	A(w)	G(v)	1248	89	0	+	0	-
CCNE1	531	T(v)	C(w)	151a	332	0	0	0	0
NPM1	572	T(w)	delT(v)	337	165	+	+	+	+
AGT	618	A(v)	C(w)	584	556	0	+	+	0
NCSTN	623	delCA(v)	CA(w)	455	515	0	0	+	+
SMUG1	628	G(w)	T(v)	770	80	+	+	0	+
PARP1	769	C(v)	T(w)	145	607	+	+	0	+
WFS1	779	A(v)	G(w)	668	253	+	+	+	+
MSX1	790	G(v)	A(w)	3649	276	+	+	+	-
EFNA1	843	A(v)	G(w)	200c	154	+	+	+	-
IL-23R	851	A(v)	C(w)	let-7e	309	0	0	+	-
RYR3	880	G(v)	A(w)	367	839	+	0	0	+
AGTR1	888	C(v)	A(w)	155	86	+	+	0	+
FGF20	903	T(v)	C(w)	433	182	-	0	0	-
DROSHA	937	C(v)	T(w)	27b	92	0	0	0	-
HOXB5	952	G(w)	A(v)	7	141	+	0	+	-
RAD51	978	A(v)	G(w)	197	718	+	+	0	-
REV3L	985	C(w)	T(v)	25	460	0	+	0	+
ORAI1	1034	T(v)	C(w)	519a	86	0	0	0	+
RAP1A	1078	C(v)	A(w)	196a	366	+	+	+	+
APP	1120	C(v)	T(w)	147	171	+	0	+	+
IL1A	1152	TTCA(v)	-(w)	122	928	0	0	0	+
CD133	1167	A(w)	C(v)	135b	667	+	+	+	+
PDCD1	1177	A(v)	G(w)	4717	889	+	0	+	0

We note that this work is an extension of the work published in [242] where a part of the dataset was included (20 instances). In [242], we found that overall, the score S_{sum} differentiates between the two alleles on 14 instances, giving better predictions than the other scores. The score S_{u} gives positive predictions for 12 instances and if S_{sum} and S_{u} are taken together for Case A, the predictions are in favor of the correct S-allele by at least one of the scores on 16 instances. We concluded that the two other scores S_{tot} and S_{P} are less sensitive to binding patterns when compared to S_{sum} and S_{u}. Therefore, we will only include these two scores in this analysis. A more detailed version of the predictions are given in Supplementary Material in Section .5

The score S_{sum} differentiates particularly well between the two alleles on 22 instances, while the S_{sum} scores are indifferent ($-1\text{kcal/mol} < \Delta S_{\text{sum}} \leq 0\text{kcal/mol}$) in 10 instances(SPI1, NCSTN, AGT, IL-23R, REV3L, ORAI1, IL1A, TYMS, CCNE1 and DROSHA). For the two other cases (HTR3E and FGF20), S_{sum} is in favor of the weaker allele (W-allele). If differences in scores in Case B are taken into account for N = 10, S_{sum} returns a strong and correct prediction for REV3L (Case A is also in favor of the S-allele, but above -1kcal/mol) and DROSHA where Case A is indifferent. These results are in line with our analysis carried out in Chapter 3 where the simple weight selection described in Section 3.4 results in a non favorable total score for HTR3E and FGF20.

The score S_{u} gives positive predictions for 20 instances and 13 indifferent predictions. However, if S_{sum} and S_{u} are taken together for Case A, the predictions are in favor of the correct S-allele by at least one of the scores on 25 instances with 2 false predictions for HTR3E and FGF20 and 7 indifferent scores.

Single allelic variations in the gene 3'UTR can affect miRNA bindings by destroying an existing miRNA-mRNA site or creating a new interaction. `RNAStrucTar` has the ability to check this SNP effect using the score S_{u} and the Case B. If S_{u} is predicted for Case A in the case of an allele and is not available with the other allele, one can conclude that the SNP created or destroyed a valid binding site. In our dataset, this was the case for the following instances: KRT81, TYMS, NPM1, AGT, WFS1, EFNA1 and AGTR1. Additionally, SNPs have the potential to interfere and change the local structures within a close proximity to the SNP position, affecting the accessibility of a binding site encapsulating the allelic

change. However, other cases showed that a SNP may change the structure in a location different from the binding region. As an example, in the case of NFKBIA and PDCD1, the difference in score S_{sum} is -4.7 and -2.3 respectively, while no sites with SNP position inside are predicted. In the case of RYR3, HOXB5 and APP, the score S_{u} coming from the SNP binding region does not contribute to the total score S_{sum}, concluding the difference in S_{sum} between the two alleles originates from the difference in accessibility within other regions of the UTR.

4.3.3 Comparison to Other Computational Methods

We compare our predictions to those produced by other accessibility-based target prediction methods, `PITA` [24] and `STarMir` [97]. For the 34 instances we consider, the `PITA` tool returns predictions in favor of the S-allele on 20 instances, with 14 indifferent scores. Thus, predictions for HTR3E, SPI1, NCSTN, AGT and IL-23R are correct by `PITA`, but not by `RNAStrucTar` (S_{sum} only). Predictions for MSLN, RAD51, NFKBIA, SMUG1, PARP1, RYR3 and AGTR1 are correct by `RNAStrucTar`, but not by `PITA`. Both tools fail on REV3L, ORAI1, IL1A, TYMS, CCNE1, FGF20 and DROSHA.

The equivalent of S_{sum} for `STarMir` predictions returns score differences in favor of the S-allele on 20 instances, with 2 indifferent outcomes (AGT and CCNE1) but 12 false predictions in favor of the W-allele on MSLN, RAD51, KRT81, MTHFD1L, MSX1, EFNA1, HOXB5, PDCD1, IL-23R, TYMS, FGF20 and DROSHA. Thus, predictions for HTR3E, SPI1, NCSTN, REV3L, ORAI1 and IL1A are correct by `STarMir`, but not by `RNAStrucTar`. Predictions for KRT81, MTHFD1L, MSX1, EFNA1, HOXB5, PDCD1, MSLN and RAD51 are correct by `RNAStrucTar`, but not by `STarMir`. Both tools fail on AGT, IL-23R, TYMS, CCNE1, FGF20 and DROSHA.

In Figure 4.4 we combine the results of the three methods. If we classify an instance as positively predicted if at least two of the methods return a prediction in favor of the S-allele, then 23 correct predictions are made. If only one positive return by a single method is required, then 30 correct predictions are produced by the three methods (only TYMS, CCNE1, FGF20 and DROSHA are rejected).

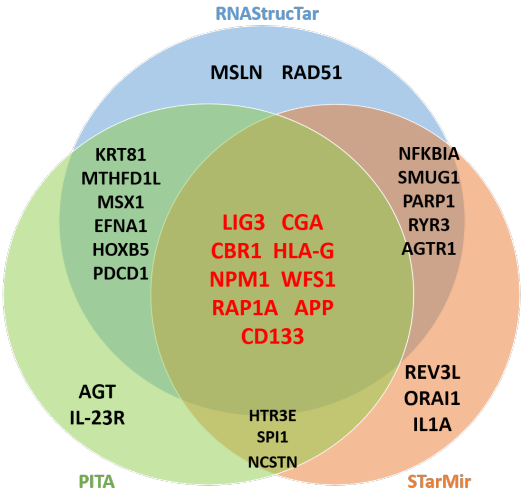


Figure 4.4: Comparison of predictions between `RNAStrucTar` and two existing methods.

4.4 RNAStrucTar: A Standalone microRNA Target Prediction Tool

The requirement to run `RNAStrucTar` using the command line may be a technical barrier for a biologist. Therefore, a user friendly version of `RNAStrucTar` with an appropriate interface has been implemented. The `RNAStrucTar` user interface is implemented using Perl tk. The user needs to input the sequences for both miRNA and the potential target and other mandatory parameters which will be needed to generate the metastable conformations: the desired energy range above the MFE and the desired energy barrier, as described in Section 4.2.1. The output results are presented to the user through a text summarizing the predictions for the different sets as described in Section 4.2.7. For each metastable conformations' set, an output file is given, detailing all predicted binding sites by `RNAStrucTar` with the corresponding energy scores, as described in Section 4.2.5. Figure 4.5 is a screenshot of the implemented user interface.

Figure 4.5: `RNAStrucTar` graphical user interface.

4.5 Conclusion

We presented in this chapter `RNAStrucTar`, a miRNA target prediction tool which incorporates target site accessibility related to metastable secondary structures close to the MFE conformation. We tested our method on 34 miRNA:mRNA interaction pairs that have been experimentally evaluated in the literature. We found that a combination of the two main scores returned by `RNAStrucTar` supports the experimental findings on 25 instances, with seven indifferent outcomes and two false classifications. If `STarMir` results (20 correct, but partly on different instances) are taken into account, then experimental findings are supported on 28 instances.

In contrast, in Chapter 2 Section 2.4 where six databases studying disease associated SNP effects on miRNA bindings have been compared, the `MicroSNiPer` and `PolymiRTS` databases successfully detected the SNPs to disturb a binding site with the associated miRNA in respectively 15 and 16 instances from the dataset. Thus, we think that `RNAStrucTar` may provide additional, useful information for miRNA target predictions, specifically in SNP related studies, where the user is studying the effects miRNA target site variations may have on the development of various human diseases.

In Table 4.2, we combine the predictions of `RNAStrucTar`, `STarMir` and `PITA` with the predictions from Chapter 2 Section 2.4. All instances are successfully predicted by at least one prediction program. Nine out of the 12 instances that are not predicted by the available databases in Chapter 2 Section 2.4 are correctly predicted by `RNAStrucTar`. Among these instances, the case [APP;T171C;miR-147] has an unknown SNP ID and therefore we were unable to include it in our analysis in Chapter 2. However, `RNAStrucTar` tool provides flexibility at the input stage by allowing the user to enter a user defined UTR and miRNA sequences with an associated SNP position and therefore we were able to correctly predict the effect of the SNP in this case.

Consequently, we believe that the proposed tool may be useful in the prediction step of the common working scheme followed in most of the comprehensive studies of miRSNPs, as described in Section 2.3.1 from Chapter 2.

Table 4.2: Summary of the predictions for the 34 instances from the dataset. SNP dbs refers to the number of databases, studying disease-associated SNP effects on miRNA bindings, correctly predicting the effect of the SNP.

	SNP dbs	RNAStrucTar	STarMir	PITA
HLA-G	6	✓	✓	✓
WFS1	5	✓	✓	✓
CD133	2	✓	✓	✓
RAP1A	1	✓	✓	✓
CGA	0	✓	✓	✓
CBR1	0	✓	✓	✓
LIG3	0	✓	✓	✓
NPM1	0	✓	✓	✓
APP	NA	✓	✓	✓
EFNA1	5	✓		✓
AGTR1	4	✓	✓	
KRT81	4	✓		✓
MSX1	4	✓		✓
SPI1	2		✓	✓
MTHFD1L	2	✓		✓
PDCD1	2	✓		✓
NCSTN	1		✓	✓
HOXB5	1	✓		✓
SMUG1	1	✓	✓	
NFKBIA	0	✓	✓	
RYR3	0	✓	✓	
PARP1	0	✓	✓	
HTR3E	4		✓	✓
RAD51	4	✓		
AGT	2			✓
REV3L	1		✓	
ORAI1	0		✓	
IL-23R	0			✓
IL1A	0		✓	
MSLN	0	✓		
CCNE1	3			
FGF20	2			
TYMS	2			
DROSHA	1			

Chapter 5

Conclusions

In this chapter, we evaluate the contributions of the work presented in this thesis and discuss its possible extensions.

5.1 Evaluation of Contribution

In this thesis, we have addressed the problem of miRNA bindings to metastable RNA secondary structures in the context of SNPs. With the aim to improve detecting the effects of the SNPs on miRNA bindings, we first looked for experimentally validated, disease associated, SNPs which have been verified to affect a miRNA binding and selected a dataset of 34 [mRNA/3'UTR; SNP; miRNA] instances from recent literature, based on strong expression level analyses. After analysis of the common working scheme followed in most of the literature sources we researched for the current work, we highlighted that the use of miRNA-target prediction methods is a crucial step in the process of miRSNP-disease association, predicting the effect of SNPs on miRNA bindings. We subsequently carried out a review of three categories of miRNA-target gene related research databases: the databases which offer repositories for experimentally validated miRNA-target interactions, the databases which provide direct access to miRNA-target predictions from several well established tools and finally databases offering SNP-effect predictions. Furthermore, we compared the results of two major computational approaches to miRNA target ranking prediction: conservation feature using `TargetScan` tool and target site accessibility feature using `PITA` and `STarMir`

tools. We conclude that using the site accessibility feature may improve miRNA-target predictions. Consequently, we studied the problem of miRNA bindings to metastable secondary structures in the context of SNPs and mRNA concentration levels i.e. whether features of miRNA bindings to metastable conformations could provide additional information supporting the differences in expression levels of the two sequences defined by a SNP. Among the different parameters we introduced and analyzed, we found that three of them, related to the average depth and average opening energy of metastable conformations, may help improve the confidence about expected miRNA-mRNA bindings in a preprocessing step of wet lab experiments. Finally, we implemented `RNAStrucTar`, a miRNA target prediction tool which incorporates target site accessibility related to metastable secondary structures close to the MFE conformation. The proposed tool outperforms similar methods when testing on the manually curated [mRNA/3'UTR; SNP; miRNA] dataset and therefore may improve miRNA-target predictions, specifically in SNP-disease association studies.

5.2 Limitations

In the literature sources we researched for the current work, the number of genes and miRNAs exposed to experimental analysis is relatively small, yet each analysis is time-consuming and costly. Additionally, the approach presented in the current work is clearly limited by the vast amount of secondary structures returned by `RNAsubopt` for longer sequences, within even a small energy range above the MFE conformation. Due to this computational limitation, only mRNA/3'UTR sequences below approximately 1100nt could be included. Because we are aiming at the applicability of `MSbind` and `RNAStrucTar` by using standard hardware facilities, we are seeking for alternative ways of identifying metastable conformations, with preference given to approximation methods, where at the cost of completeness (within an acceptable error range) metastable conformations are calculated within a feasible time. Moreover, the proposed `RNAStrucTar` program is currently accessible by a command-line interface, that needs to be installed in Linux and therefore non specialist users may find it difficult to perform runs

using this implemented user interface.

5.3 Future Research

In view of extending our work on miRNA-target prediction using metastable secondary structures, we intend to make `RNAStrucTar` a web service for remote calls, thus eliminating the need for a local installation. An interesting extension of our current work may include a similar study on target prediction in the context of RNA-binding proteins (RBPs). In fact, the 3'UTR is also a favoured target for RBPs that regulate maturation, stability, transfer, localization and translation of mRNAs [246]. It has been also shown that the mRNA accessibility of a potential RBP target site plays an important role in finding whether the RBP actually binds to the site [247; 248]. Therefore, we intend to incorporate target site accessibility related to metastable secondary structures, representing local minima and close to the MFE conformation, into predictions of target selection by RBPs. For future research, we would like to investigate haplotypes and miRNA bindings, as discussed in Martin *et al.* [179], and the analysis of SNPs in miRNAs, see [178]. Furthermore, we plan to examine data provided by eQTL studies as an additional source of experimental data bearing information about the effect of SNPs on miRNA bindings to putative targets. For example, Westra *et al.* [249] report that trans-eQTL SNPs are enriched within miRNA binding sites in blood cell lines studied in the article, see Figure 1 in Westra *et al.* [249]. Although in the present study we focus on miRNA binding within 3'UTRs, we also plan to look into the case where miRNA binding sites near the terminal coding region along with the presence of SNPs may affect the overall 2D or 3D structure.

Software Implementation and Availability

For a given set of metastable secondary structures, the application `MSBind` used in Chapter 3 is available at http://kks.inf.kcl.ac.uk/MSbind.html for calculating features of metastable conformations determined by putative miRNA binding sites.

The `RNAStrucTar` program described in Chapter 4 is written in C++/Perl. The sourcecode and the disease-associated SNPs within miRNA target sites dataset are currently available on request. The current version can be downloaded in a compressed form tar.gz format. In Linux, it can be decompressed using *tar -zxvf rnastructar.tar.gz*. `RNAStrucTar` is using the Vienna RNA package [237] that needs to be compiled (*cd ViennaRNA*.

Supplementary Tables

.1 Reference numbers and sequences from the NCBI, the dbSNP and the Ensembl databases

This document completes the information about the 34 cases included in our dataset (Chapter 2). For each case, we provide the information related to the UTR sequence used for testing in the related publication ('In publication' section). 'In dbSNP' section includes the information retrieved related to the SNP ID, NCBI's dbSNP official SNP identifiers ("rs-numbers"). 'In Ensembl BioMart' section lists the UTR sequences retrieved from the BioMart from the Ensembl database, using both the RefSeq IDs from the previous sections and the associated gene name for other UTR sequences. The SNP position is indicated in each sequence for each case.

LIG3

L(3'UTR) nt	124
mRNA RefSeq No	NM 002311.4
Ensembl Gene ID	ENSG00000005156
Ensembl Transcript ID	ENST00000262327
SNP ID	rs4796030

In publication

LIG3 rs4796030 C>A and the SNP is located within a predicted miRNA-binding site.

In dbSNP

TCAAAATTTACATTAAAGGGAAAAG[A/C]CCAGTCTGGGTGTGGGAATGCAGCA
NM_002311.4:c.*83A>C
NM_013975.3:c.2796+342A>C

In Ensembl Biomart

NM_002311.4

>ENSG00000005156|ENST00000262327|LIG3
AACAGCCCGGCCTAGCCAGGAGAGACTGCAGGGACTCACTCAGCTGCTGGCCCCAAGTCAAAATTTACATTAAAGGGAAAAG[A/C]CCAGTCT
GGGTGTGGGAATGCAGCATTGAGTTTGTGGTCAG

NM_013975.3

>ENSG00000005156|ENST00000378526|LIG3 Sequence of length 5237
SNP not found in the sequence

Other available LIG3 3'UTR transcripts in Biomart:

> ENSG00000005156|ENST00000588109|LIG3 Sequence of length 477

Average Sequence length: 1946

Unavailable Sequences : ENST00000588713, ENST00000592244, ENST00000586119, ENST00000585740, ENST00000586435, ENSG00000005156, ENST00000590181, ENST00000590630.

MSLN

L(3'UTR) nt	132
mRNA RefSeq No	NM_005823
Ensembl Gene ID	ENST00000566549
Ensembl Transcript ID	ENSG00000102854

SNP ID rs1057147

In publication

The RefSeq of the used 3'UTR of *MSLN* is given (NM_005823.5). The SNP rs1057147 G > A falls within the seed match of the given binding site (binding region is showed in a figure).

In dbSNP

GCCTGGCCAGGAGCAGGCACGGGTG[A/G]TCCCCGTTCCACCCCAAGAGAACTC
NM_001177355.1:c.*69G>A
NM_005823.5:c.*69G>A
NM_013404.4:c.*69G>A

In Ensembl Biomart

NM_005823

>ENSG00000102854|ENST00000545450 Sequence of length 125
GGGCCCCACTCCCTTGCTGGCCCCAGCCCTGCTGGGGATCCCCGCCTGGCCAGGAGCAGGCACGGGTG[A/G]TCCCCGTTCC
ACCCCAAGAGAACTCGCGCTCAGTAAACGGGAACATGCCCCCTGCA

NM_001177355

>ENSG00000102854|ENST00000563941 Sequence of length 129

NM_013404

>ENSG00000102854|ENST00000382862 Sequence of length 128

Other available 3'UTR transcripts in Biomart:

>ENSG00000102854|ENST00000566549 Sequence of length 132
>ENSG00000102854|ENST00000566269 Sequence of length 65

Average Sequence length: 116

Unavailable Sequences : ENST00000563651, ENST00000561896, ENST00000569566, ENST00000569566.

CGA

L(3'UTR) nt	258
mRNA RefSeq No	NM_000735
Ensembl Gene ID	ENSG00000135346
Ensembl Transcript ID	ENST00000627148
SNP ID	rs6631

In publication

The SNP position falls within the given predicted binding site (Substitution of T by A at rs6631, the binding region is showed in a figure).

In dbSNP

TTTCATTTGGAATGAATACAGCATT[A/T]AGCTTGTTCCACTGCAAATAAAGCC
NM_000735.3:c.*215T>A
NM_001252383.1:c.*215T>A

In Ensembl Biomart

NM_000735

>ENSG00000135346|ENST00000627148
ATGTTTTACCAAGTGCTGTCTTGATGACTGCTGATTTTCTGGAATGGAAAATTAAGTTGTTTAGTGTTTATGGCTTTGTGAGATAAAACTCTCC
TTTTCCTTACCATACCACTTTGACACGCTTCAAGGATATACTGCAGCTTTACTGCCTTCCTCCTTATCCTACAGTACAATCAGCAGTCTAGTTC
TTTTCATTTGGAATGAATACAGCATT[A/T]AGCTTGTTCCACTGCAAATAAAGCCTTTTAAATCATCATTCAA

NM_001252383

>ENSG00000135346|ENST00000610310 Sequence of length 72
SNP not found in the sequence.

Other available 3'UTR transcripts in Biomart:
>ENSG00000135346|ENST00000369582 Sequence of length 253
>ENSG00000135346|ENST00000630630 Sequence of length 427

Average Sequence length: 252

Unavailable Sequences : ENST00000627552, ENST00000625577.

CBR1

L(3'UTR) nt	284
mRNA RefSeq No	NM 001757.2
Ensembl Gene ID	ENSG00000159228
Ensembl Transcript ID	ENST00000290349
SNP ID	rs9024

In publication

The GenBank sequence NM_001757.2 is used. The binding site is in direct proximity to the SNP position 133 (rs9024, *CBR1* 1096G>A , the binding region is showed in a figure).

In dbSNP

CTCTTATCAATTAGCACTCACTAAT[A/G]TACTACTAATTGAGCAACCTACGCA
NM_001286789.1:c.*1076G>A
NM_001757.3:c.*133G>A

In Ensembl Biomart

NM_001286789

>ENSG00000159228|ENST00000530908|CBR1 Sequence of length 1227

NM_001757

>ENSG00000159228|ENST00000290349|CBR1
GCTGGGCTCACAGCTCCATCCATGGGCCCCATTTTGTACCTTGTCCTGAGTTGGTCCAAAGGGCATTTACAATGTCATAAATATCCTTATATAA
GAAAAAAAATGATCTCTTATCAATTAGCACTCACTAAT[A/G]TACTACTAATTGAGCAACCTACGCACTCAGTTGACTACGTAAATCTGTC
AGGTCTTTTGTGATTTCCTCTGATGCAGGAGAGGAAAAATTGTAATTGATGAAAATAATGAATGAAAATCAACAGATGAATAAATGGTTCTTTA
TAAGTGTC

Other available 3'UTR transcripts in Biomart:

>ENSG00000159228|ENST00000439427|CBR1 Sequence of length 386
>ENSG00000159228|ENST00000399191|CBR1 Sequence of length 88

Average Sequence length: 496

Unavailable Sequences : ENST00000466328.

HTR3E

L(3'UTR) nt	302
mRNA RefSeq No	NM_182589.2
Ensembl Gene ID	ENSG00000186038
Ensembl Transcript ID	ENST00000335304
SNP ID	rs56109847

In publication

The SNP position falls within the given miRNA binding site (*HTR3E* binding region is showed in a figure, where the variant c.*76G>A (rs62625044) disrupts the perfect seed match).

In dbSNP

rs62625044 has merged into rs56109847

ACTGGCCAGGTCTCCCCCCTTTCCT[A/G]AGTACCAACTATCATATCCCCAAAG
NM_001256613.1:c.*76G>A
NM_001256614.1:c.*76G>A
NM_182589.2:c.*76G>A
NM_198313.2:c.*76G>A
NM_198314.2:c.*76G>A

In Ensembl Biomart

NM_182589

>ENSG00000186038|ENST00000335304|HTR3E
GCAGGTGCTCACCTGCCAACTTCAGTCTGGAGCTTCTCTTGCCTCCAGGGACTGGCCAGGTCTCCCCCCTTTCCT[A/G]AGTACCAACTAT
CATATCCCCAAAGATGACTGAGTCTCTGCTGTATTCCATGTATCCCAATCCGGTCCTGCTGATCAATTCCAATCCCAGACATTTCTCCCTGTTC
CTGCATTTTGTTGGCTTCCTTCAGTCCTACCATATGGTTCTAGGTCCCTCTTACGTCATCTGCATAGCAGACTATACCTCTTCTGTCCGCTGAC
TTGCCCAATAAATAATTCTGCAGAGA

NM_001256614

>ENSG00000186038|ENST00000440596 Sequence unavailable

NM_001256613

>ENSG00000186038|ENST00000415389|HTR3E Same sequence as NM_182589
GCAGGTGCTCACCTGCCAACTTCAGTCTGGAGCTTCTCTTGCCTCCAGGGACTGGCCAGGTCTCCCCCCTTTCCT[A/G]AGTACCAACTATC
ATATCCCCAAAGATGACTGAGTCTCTGCTGTATTCCATGTATCCCAATCCGGTCCTGCTGATCAATTCCAATCCCAGACATTTCTCCCTGTTCC
TGCATTTTGTTGGCTTCCTTCAGTCCTACCATATGGTTCTAGGTCCCTCTTACGTCATCTGCATAGCAGACTATACCTCTTCTGTCCGCTGACT
TGCCCAATAAATAATTCTGCAGAGA

NM_198313

>ENSG00000186038|ENST00000425359|HTR3E Sequence unavailable

NM_198314

>ENSG00000186038|ENST00000436361|HTR3E Sequence unavailable

No other available 3'UTR transcripts in Biomart.

Average Sequence length: 302

Unavailable Sequences : ENST00000431041

KRT81

L(3'UTR) nt	342
mRNA RefSeq No	NM_002281
Ensembl Gene ID	ENSG00000205426
Ensembl Transcript ID	ENST00000327741
SNP ID	rs3660

In publication

rs3660G>C: the SNP position falls within a binding site for the studied miRNAs (The rs3660 G-to-C change produces the lost of a nucleotide binding in the seed sequence region, showed in a figure).

In dbSNP

GAGCAAGTGCTCAGCTACTTCTCCT[C/G]CACTTTGAAAGACCCCTCCCACTCC
NM_001320198.1:c.-5+10207C>G
NM_002281.3:c.*102G>C

In Ensembl Biomart

NM_001320198 (Gene KRT86 and not KRT81)

>ENSG00000170442|ENST00000423955|KRT86 Sequence of length 582

NM_002281

>ENSG00000205426|ENST00000327741|KRT81
GCACCCCAACTCAAGTCCCAGGCCCCAGGCATCTTTCCTGCCCTGCCTTGCTTGGCCCATCCAGTCCAGGCGCCTGGAGCAAGTGCTCAGCTAC
TTCTCCT[C/G]CACTTTGAAAGACCCCTCCCACTCCTGGCCTCACATTTCTCTGTGTGATCCCCCACTTCTGGGCTCTGCCACCCCACAGT
GGGAAAGGCCACCCTAGAAAGAAGTCCGCTGGCACCCATAGGAAGGGGCCTCAGGAGCAGGAAGGGCCAGGACCAGAACCTTGCCCACGGCAAC
TGCCTTCCTGCCTCTCCCCTTCCTCCTCTGCTCTTGATCTGTGTTTCAATAAATTAATGTAGCCAA

Average Sequence length: 342

Unavailable Sequences : ENST00000615839.

SPI1

L(3'UTR) nt	369
mRNA RefSeq No	NM_003120
Ensembl Gene ID	ENSG00000066336
Ensembl Transcript ID	ENST00000378538
SNP ID	rs1057233

In publication

The SNP rs1057233 C/T alters the target sequence of the miRNA (The binding site is showed in a figure where the SNP generates a mismatch in the target site of the seed region).

In dbSNP

CCCTCTCACCCCTCCCACCCATTAA[C/T]CTCCTCCCAAAAAACAAGTAAAGTT
NM_001080547.1:c.*330C>T
NM_003120.2:c.*330C>T

In Ensembl Biomart

NM_001080547

>ENSG00000066336|ENST00000227163|SPI1 Sequence of length 314
SNP not found in the sequence

NM_003120

>ENSG00000066336|ENST00000378538|SPI1
GCCCGCAGCCCCCGCCGGGCCCCGCCAGGCCTCCCCGCTGGCCATAGCATTAAGCCCTCGCCCGGCCCGGACACAGGGAGGACGCTCCCGGGGC
CCAGAGGCAGGACTGTGGCGGGCCGGGCCTCGCCTCACCCGCCCCCTCCCCCCACTCCAGGCCCCCTCCACATCCCGCTTCGCCTCCCTCCAGG
ACTCCACCCCGGCTCCCGGACGCCAGCTGGGCGTCAGACCCCACCGGGGCAACCTTGCAGAGGACGACCCGGGGTACTGCCTTGGGAGTCTCAA
GTCCGTATGTAAATCAGATCTCCCCTCTCACCCCTCCCACCCATTAA[C/T]CTCCTCCCAAAAAACAAGTAAAGTTATTCTCAATCCA

Other available 3'UTR transcripts in Biomart:

>ENSG00000066336|ENST00000533030|SPI1 Sequence of length 383
>ENSG00000066336|ENST00000533968|SPI1 Sequence of length 418

Average Sequence length: 370

HLA-G

L(3'UTR) nt	386
mRNA RefSeq No	NM 002127.5
Ensembl Gene ID	ENSG00000204632
Ensembl Transcript ID	ENST00000360323
SNP ID	rs1063320

In publication

The SNP *HLA-G* +3142C/G falls within the seed match of the target site, showed in a figure.

In dbSNP

atgtctccgtctctgtctcaaatttgtggt[c/g]cactgagctataacttacttctgtattaaa
NM_002127.5:c.*233C>G

In Ensembl Biomart

NM_002127

>ENSG00000204632|ENST00000360323|HLA-G
AAAGGAGGGAGCTACTCTCAGGCTGCAATGTGAAACAGCTGCCCTGTGTGGGACTGAGTGGCAAGTCCCTTTGTGACTTCAAGAACCCTGACTC
CTCTTTGTGCAGAGACCAGCCCACCCCTGTGCCCACCATGACCCTCTTCCTCATGCTGAACTGCATTCCTTCCCCAATCACCTTTCCTGTTCCA
GAAAAGGGGCTGGGATGTCTCCGTCTCTGTCTCAAATTTGTGGT[c/g]CACTGAGCTATAACTTACTTCTGTATTAAAATTAGAATCTGAG
TATAAATTTACTTTTTCAAATTATTTCCAAGAGAGATTGATGGGTTAATTAAAGGAGAAGATTCCTGAAATTTGAGAGACAAAATAAATGGAAG
ACATGAGAACTTTCCA

Several available sequences are available in the Ensembl Biomart database. The length of the sequences range between 383 and 389 nucleotides.

MTHFD1L

L(3'UTR) nt	393
mRNA RefSeq No	NM_001242767

Ensembl Gene ID	ENSG00000120254
Ensembl Transcript ID	ENST00000611279
SNP ID	rs7646

In publication

The NCBI Reference of the used sequence is given (NM_001242767 of length 393). A detailed effect on the binding region is given: the SNP rs7646 is present in position 120, the seed sequence of miR-197 is in position 121–127.

In dbSNP

AAGAAATCTTTGGCCTATTTCACCA[A/C/T]GTCTCCAGCATTGCATAACAGACAT

NM_001242767.1:c.*120A>G
NM_001242768.1:c.*120A>G
NM_015440.4:c.*120A>G

In Ensembl Biomart

NM_001242767

>ENSG00000120254|ENST00000611279|MTHFD1L
GTGGACAAGGCTCTCACAGGACCCGATGCAGACTCCTGAAACAGACTACTCTTTGCCTTTTTGCTGCAGTTGGAGAAGAAACTGAATTTGAAAA
ATGTCTGTTATGCAATGCTGGAGAC[A/G]TGGTGAAATAGGCCAAAGATTTCTTCTTCGTTCAAGATGAATTCTGTTCACAGTGGAGTATGGT
GTTCGGCAAAAGGACCTCCACCAAGACTGAAAGAAACTAATTTATTTCTGTTTCTGTGGAGTTTCCATTATTTCTACTGCTTACACTTTAGAAT
GTTTATTTTATGGGGACTAAGGGATTAGGAGTGTGAACTAAAAGGTAACATTTTCCACTCTCAAGTTTTCTACTTTGTCTTTGAACTGAAAATA
AACATGGATCTAGAAAACCAA

NM_001242768

>ENSG00000120254|ENST00000618312|MTHFD1L Sequence of length 391

NM_015440

>ENSG00000120254|ENST00000367321|MTHFD1L Sequence of length 393

Other available 3'UTR transcripts in Biomart:

>ENSG00000120254|ENST00000453602|MTHFD1L Sequence of length 224
>ENSG00000120254|ENST00000367307|MTHFD1L Sequence of length 114
>ENSG00000120254|ENST00000441122|MTHFD1L Sequence of length 574
>ENSG00000120254|ENST00000420192|MTHFD1L Sequence of length 315
>ENSG00000120254|ENST00000450635|MTHFD1L Sequence of length 253

Average Sequence length: 261

Unavailable Sequences :

Sequence unavailable ENST00000423867, ENST00000478643, ENST00000421497, ENST00000367308.

NFKBIA

L(3'UTR) nt	502
mRNA RefSeq No	NM_020529
Ensembl Gene ID	ENSG00000100906
Ensembl Transcript ID	ENST00000216797

SNP ID rs696

In publication

The SNP (2758A>G (rs696) in NFκBIA) falls within the given binding site (the binding site is showed in a figure, where the G variant of NFkBIA generates a potential perfect seed site for miR-449a).

In dbSNP

CTTATATCCACACTGCACACTGCCT[A/G]GCCCAAAACGTCTTATTGTGGTAGG
NM_020529.2:c.*126G>A

In Ensembl Biomart

NM_020529

>ENSG00000100906|ENST00000216797|NFKBIA
GCGCAAAGGGGCTGAAAGAACATGGACTTGTATATTTGTACAAAAAAAAAGTTTTATTTTTCTAAAAAAAGAAAAAAGAAGAAAAAATTTAAAG
GGTGTACTTATATCCACACTGCACACTGCCT[A/G]GCCCAAAACGTCTTATTGTGGTAGGATCAGCCCTCATTTTGTTGCTTTTGTGAACTTT
TTGTAGGGGACGAGAAAGATCATTGAAATTCTGAGAAAACTTCTTTTAAACCTCACCTTTGTGGGGTTTTTGGAGAAGGTTATCAAAAATTTCA
TGGAAGGACCACATTTTATATTTATTGTGCTTCGAGTGACTGACCCCAGTGGTATCCTGTGACATGTAACAGCCAGGAGTGTTAAGCGTTCAGT
GATGTGGGGTGAAAAGTTACTACCTGTCAAGGTTTGTGTTACCCTCCTGTAAATGGTGTACATAATGTATTGTTGGTAATTATTTTGGTACTTT
TATGATGTATATTTATTAAACAGATTTTTACAAATG

Other available 3'UTR transcripts in Biomart:

>ENSG00000100906|ENST00000554001|NFKBIA Sequence of length 1075
>ENSG00000100906|ENST00000557389|NFKBIA Sequence of length 468
>ENSG00000100906|ENST00000557140|NFKBIA Sequence of length 473

Average Sequence length: 629

Unavailable Sequences : ENST00000553342, ENST00000557100, ENST00000555629, ENST00000555371, ENST00000556664, ENST00000557459.

TYMS

L(3'UTR) nt	502
mRNA RefSeq No	NM_001071
Ensembl Gene ID	ENSG00000176890
Ensembl Transcript ID	ENST00000323274
SNP ID	rs2790

In publication

The rs2790 A>G falls within the seed region of the given target site (showed in a figure). The length of the used 3'UTR sequence is given.

In dbSNP

AAAAGTTCTTTTTGCTCTAAAAGAA[A/G/T]AAGGAACTAGGTCAAAAATCTGTCC
NM_001071.2:c.*89A>G
NM_001126123.3:c.*145-10T>C
NM_001318759.1:c.*145-10T>C
NM_017512.5:c.*1219T>C
NM_202758.3:c.*1219T>C

In Ensembl Biomart

NM_001071

>ENSG00000176890|ENST00000323274|TYMS

GGTGCTTTCAAAGGAGCTCGAAGGATATTGTCAGTCTTTAGGGGTTGGGCTGGATGCCGAGGTAAAAGTTCTTTTTGCTCTAAAAGAA[A/G/T]AAGGAACTA
GGTCAAAAATCTGTCCGTGACCTATCAGTTATTAATTTTTAAGGATGTTGCCACTGGCAAATGTAACTGTGCCAGTTCTTTCCATAATAAAAGGCTTTGAGTTAACT
CACTGAGGGTATCTGACAATGCTGAGGTTATGAACAAAGTGAGGAGAATGAAATGTATGTGCTCTTAGCAAAAACATGTATGTGCATTTCAATCCCACGTACTTATA
AAGAAGGTTGGTGAATTTCACAAGCTATTTTTGGAATATTTTTAGAATATTTTAAGAATTTCACAAGCTATTCCCTCAAATCTGAGGGAGCTGAGTAACACCATCGA
TCATGATGTAGAGTGTGGTTATGAACTTTAAAGTTATAGTTGTTTTATATGTTGCTATAATAAAGAAGTGTTCTGCATTCGTCCACGCTTTGTTCATTCTGTACTGC
CACTTATCTGCTCAGTTCCTTCCTAAAATAGATTAAAGAACTCTCCTTAAGTAAA

NM_001126123 **(Gene ENOSF1)**
>ENSG00000132199|ENST00000383578|ENOSF1
NM_001318759 **(Gene ENOSF1)**
>ENSG00000132199|ENST00000383578|ENOSF1
NM_017512 **(Gene ENOSF1)**
>ENSG00000132199|ENST00000251101|ENOSF1
NM_202758 **(Gene ENOSF1)**

>ENSG00000132199|ENST00000340116|ENOSF1

Other available 3'UTR transcripts in Biomart:

>ENSG00000176890|ENST00000323224|TYMS Sequence of length 487
>ENSG00000176890|ENST00000323274|TYMS Sequence of length 581

Average Sequence length: 523

Unavailable Sequences : ENST00000581920, ENST00000579128, ENST00000323250, ENST00000584122.

CCNE1

L(3'UTR) nt	531
mRNA RefSeq No	NM_001238
Ensembl Gene ID	ENSG00000105173
Ensembl Transcript ID	ENST00000262643
SNP ID	rs3218073

In publication

CCNE1 rs3218073 polymorphism located at miRNA-151 binding site (Wild type binding region is howed: GTGGCT**C**TCCTCG).

In dbSNP

TGTGGAGGGCCACGGTGGCGTGGCT[C/T]TCCTCGCAGGTGTTCTGGGCTCCGT
NM_001238.2:c.*332C>T

In Ensembl Biomart

NM_001238

>ENSG00000105173|ENST00000262643|CCNE1
CCACCCCATCCTTCTCCACCAAAGACAGTTGCGCGCCTGCTCCACGTTCTCTTCTGTCTGTTGCAGCGGAGGCGTGCGTTTGCTTTTACAGATA
TCTGAATGGAAGAGTGTTTCTTCCACAACAGAAGTATTTCTGTGGATGGCATCAAACAGGGCAAAGTGTTTTTTATTGAATGCTTATAGGTTTT
TTTTAAATAAGTGGGTCAAGTACACCAGCCACCTCCAGACACCAGTGCGTGCTCCCGATGCTGCTATGGAAGGTGCTACTTGACCTAAGGGACT
CCCACAACAACAAAAGCTTGAAGCTGTGGAGGGCCACGGTGGCGTGGCT[C/T]TCCTCGCAGGTGTTCTGGGCTCCGTTGTACCAAGTGGAGC
AGGTGGTTGCGGGCAAGCGTTGTGCAGAGCCCATAGCCAGCTGGGCAGGGGGCTGCCCTCTCCACATTATCAGTTGACAGTGTACAATGCCTTT
GATGAACTGTTTTGTAAGTGCTGCTATATCTATCCATTTTTTAATAAAGATAATACTGTTTTTGA

Other available 3'UTR transcripts in Biomart:

>ENSG00000105173|ENST00000576532|CCNE1 Sequence of length 233
>ENSG00000105173|ENST00000444983|CCNE1 Sequence of length 435
>ENSG00000105173|ENST00000357943|CCNE1 Sequence of length 531

Average Sequence length: **432**

Unavailable Sequences : ENST00000586912, ENST00000574121, ENST00000575243

NPM1

L(3'UTR) nt	572
mRNA RefSeq No	NM_002520
Ensembl Gene ID	ENSG00000181163
Ensembl Transcript ID	ENST00000296930
SNP ID	rs34351976

In publication

The delT polymorphism in *NPM1* 3'-UTR falls within the seed region of the showed target site (AAAATGCCTGTTTAG)

In dbSNP

CCAAGAATGTGTTGTCCAAAATGCC[-/T]GTTTAGTTTTTAAAGATGGAACTCC
NM_002520.6:c.*165delT
NM_199185.3:c.*165delT

In Ensembl Biomart

NM_002520

>ENSG00000181163|ENST00000296930|NPM1
GAAAATAGTTTAAACAATTTGTTAAAAAATTTTCCGTCTTATTTCATTTCTGTAACAGTTGATATCTGGCTGTCCTTTTTATAATGCAGAGTGA
GAACTTTCCCTACCGTGTTTGATAAATGTTGTCCAGGTTCTATTGCCAAGAATGTGTTGTCCAAAATGCC[-
/T]GTTTAGTTTTTAAAGATGGAACTCCACCCTTTGCTTGGTTTTAAGTATGTATGGAATGTTATGATAGGACATAGTAGTAGCGGTGGTCAGA
CATGGAAATGGTGGGGAGACAAAAATATACATGTGAAATAAAACTCAGTATTTTAATAAAGTAGCACGGTTTCTATTGACTTATTTAACTGCTT
TATACTTTGTCAAAGAAATAATTAATGTAGTTAGGAATGGCAAATAGTCTTGTAAAATTCTATGAGAATGTCCCTGCCCTCCCCTTCAATATTC
TCTCTGGAGCTAACCACTTTTTCATCATAAGGATTTAGTGCTGTGTTCCCACCTCCTGATGATAGTTAACAATTATTATAACTATGCAACATGT
TTCCAAATGTTCCATTAGACCTCCTATCTGCCTA

NM_199185

>ENSG00000181163|ENST00000351986|NPM1 Sequence of length 319

Other available 3'UTR transcripts in Biomart:

>ENSG00000181163|ENST00000517671|NPM1 Sequence of length 318
>ENSG00000181163|ENST00000393820|NPM1 Sequence of length 720

Average Sequence length: **482**

Unavailable Sequences : ENST00000521260, ENST00000524204, ENST00000523622, ENST00000518587, ENST00000521672, ENST00000523339, ENST00000521710, ENST00000519955.

AGT

L(3'UTR) nt	618
mRNA RefSeq No	NM_000029
Ensembl Gene ID	ENSG00000135744
Ensembl Transcript ID	ENST00000366667
SNP ID	rs7079

In publication

The hAGT +11525 C/A polymorphism (rs7079) occurs in the seed binding sequence of miR-584 and miR-31 (binding region is showed in a figure).

In dbSNP

CTGAATTTCTGTTTGAATGCGGAAC[A/C]ATAGCTGGTTATTTCTCCCTTGTGT

NM_000029.3:c.*556C>A

In Ensembl Biomart

NM_000029

>ENSG00000135744|ENST00000366667|AGT
GGCCAGGGCCCCAGAACACAGTGCCTGGCAAGGCCTCTGCCCCTGGCCTTTGAGGCAAAGGCCAGCAGCAGATAACAACCCCGGACAAATCAGC
GATGTGTCACCCCCAGTCTCCCACCTTTTCTTCTAATGAGTCGACTTTGAGCTGGAAAGCAGCCGTTTCTCCTTGGTCTAAGTGTGCTGCATGG
AGTGAGCAGTAGAAGCCTGCAGCGGCACAAATGCACCTCCCAGTTTGCTGGGTTTATTTTAGAGAATGGGGGTGGGGAGGCAAGAACCAGTGTT
TAGCGCGGGACTACTGTTCCAAAAAGAATTCCAACCGACCAGCTTGTTTGTGAAACAAAAAAGTGTTCCCTTTTCAAGTTGAGAACAAAAATTG
GGTTTTAAAATTAAAGTATACATTTTTGCATTGCCTTCGGTTTGTATTTAGTGTCTTGAATGTAAGAACATGACCTCCGTGTAGTGTCTGTAAT
ACCTTAGTTTTTTCCACAGATGCTTGTGATTTTTGAACAATACGTGAAAGATGCAAGCACCTGAATTTCTGTTTGAATGCGGAAC[A/C]AT
AGCTGGTTATTTCTCCCTTGTGTTAGTAATAAACGTCTTGCCACAATAAGCCTCCAAAAA

No other available 3'UTR transcripts in Biomart

NCSTN

L(3'UTR) nt	681
mRNA RefSeq No	NM_001290184
Ensembl Gene ID	ENSG00000162736
Ensembl Transcript ID	ENST00000368063
SNP ID	rs141849450

In publication

The delCA515–516 (*rs141849450*) variant falls within the seed region of the given target site (a figure shows the delCA515–516 disrupts a perfect seed match) . The full length (~700 bp) of the used 3'UTR of human NCSTN.

In dbSNP

GTCCTTTCTCCAGGCCCTCAGATGG[-/CA]CATTAGGGTGGGCGTGCTGCGGGTG

NM_001290184.1:c.*515_*516delCA

NM_001290186.1:c.*515_*516delCA

NM_015331.2:c.*515_*516delCA

In Ensembl Biomart

NM_001290184

>ENSG00000162736|ENST00000368063|NCSTN

GGAGGACCCCAGCTTTTCTTGCCAGCTCAGCAGTTCACTTCCTAGAGCATCTGTCCCACTGGGACACAACCACTAATTTGTCACTGGAACCTCC
CTGGGCCTGTCTCAGATTGGGATTAACATAAAAGAGTGGAACTATCCAAAAGAGACAGGGAGAAATAAATAAATTGCCTCCCTTCCTCCGCTCC
CCTTTCCCATCACCCCTTCCCCATTTCCTCTTCCTTCTCTACTCATGCCAGATTTTGGGATTACAAATAGAAGCTTCTTGCTCCTGTTTAACTC
CCTAGTTACCCACCCTAATTTGCCCTTCAGGACCCTTCTACTTTTTCCTTCCTGCCCTGTACCTCTCTCTGCTCCTCACCCCCACCCCTGTACC
CAGCCACCTTCCTGACTGGGAAGGACATAAAAGGTTTAATGTCAGGGTCAAACTACATTGAGCCCCTGAGGACAGGGGCATCTCTGGGCTGAGC
CTACTGTCTCCTTCCCACTGTCCTTTCTCCAGGCCCTCAGATGG[-/CA]CATTAGGGTGGGCGTGCTGCGGGTGGGTATCCCACCTCCAG
CCCACAGTGCTCAGTTGTACTTTTTATTAAGCTGTAATATCTATTTTTGTTTTTGTCTTTTTCCTTTATTCTTTTTGTAAATATATATATAATG
AGTTTCATTAAAATAGATTATCCCACACGA

NM_001290186

>ENSG00000162736|ENST00000294785|NCSTN Same sequence of length 681

NM_015331

>ENSG00000162736|ENST00000294785|NCSTN Same sequence of length 681

Average Sequence length: 681

Unavailable Sequences : ENST00000469159, ENST00000491332, ENST00000424754, ENST00000421914, ENST00000424645, ENST00000435149, ENST00000438008, ENST00000491390, ENST00000467837, ENST00000459963, ENST00000437169, ENST00000465223

SMUG1

L(3'UTR) nt	628
mRNA RefSeq No	NM_001243787
Ensembl Gene ID	ENSG00000123415
Ensembl Transcript ID	ENST00000337581
SNP ID	rs2233921

In publication

The SMUG1 rs2233921 (G/T) is in position 80 and is within target-binding sites. The authors mention the use of a 640 bp long 3'UTR in the article.

In dbSNP

CAAGCAGATGACAACACATCTCCTG[G/T]ACTGGAGCAAAAGGTCCTTCTGTGC
NM_001243787.1:c.*80G>T

NM_001243788.1:c.*80G>T

NM_001243789.1:c.406-369G>T

NM_001243790.1:c.406-369G>T

NM_001243791.1:c.406-369G>T

NM_014311.2:c.*80G>T

In Ensembl Biomart

NM_001243787

>ENSG00000123415|ENST00000337581|SMUG1
GTGCCCTTGGGGCCTTGCATGGGACACATTCAAGACCTCGAAGTCATTCTTGGCCAAGCAGATGACAACACATCTCCTG[G/T]ACTGGAGCA
AAAGGTCCTTCTGTGCACCCTGGTCGCTGGGAAACGTATTCTTTGATCTGTTGAACTGTCTTCCAACCTGCCATGGCAGTTTTGACACTACTCC
TGTTTGCCCTCCTGATTCCTGCTTTCTTTACCTTTTAACATTGCCCCTTTCAGGGGACCCCACTTTGTAGGGAATCTGCAGAAGGTGTGCTTTT
GCACTTGCAGACTGCTCTACCTCAGTGTTTCCTTGGGAGACTTTATTCAGCTGAGAGTGCCCTAGACAGTAACTTCTAAGGTCACGTTTACTAT

TTCAGAGGAAATATCTTGCCAGGATACCTACCCATCCTTATAGAACAGTTACCTTTAGCTGACCCCTTTCCTCACAGGGACCAAGACAAAGCAT
GGGACATGAAATTAAGAGTGAACTTCTTATGGGAGGCTGCAGCTGGATCAGAGGAAAAATCCAGTGTGACAGAGTGCAAGTCAGAAGACCTGGC
TTTTCATCCCAGCTTTGAAACTTGGAACTTTTTGATTGACAAATTAATAAACCTCTCTATGCCTCAGGC

NM_001243788

>ENSG00000123415|ENST00000508394|SMUG1 Sequence of length 821

NM_001243789

>ENSG00000123415|ENST00000513838|SMUG1 Sequence of length 1161

NM_001243790

>ENSG00000123415|ENST00000514685|SMUG1 Sequence of length 66

NM_001243791

>ENSG00000123415|ENST00000243112|SMUG1 Sequence of length 48

NM_014311

>ENSG00000123415|ENST00000337581|SMUG1 Same sequence of length 628

Other available 3'UTR transcripts in Biomart:

>ENSG00000123415|ENST00000401977|SMUG1 Sequence of length 142
>ENSG00000123415|ENST00000505128|SMUG1 Sequence of length 2103
>ENSG00000123415|ENST00000506595|SMUG1 Sequence of length 66
>ENSG00000123415|ENST00000514196|SMUG1 Sequence of length 24
>ENSG00000123415|ENST00000509864|SMUG1 Sequence of length 190
>ENSG00000123415|ENST00000514685|SMUG1 Sequence of length 66
>ENSG00000123415|ENST00000635546|SMUG1 Sequence of length 465

Average Sequence length: 478

Unavailable Sequences : ENST00000503306, ENST00000511522, ENST00000504797, ENST00000506169, ENST00000505662, ENST00000507904, ENST00000634429, ENST00000635234, ENST00000503231, ENST00000504338, ENST00000509078, ENST00000503447, ENST00000511854, ENST00000505597.

PARP1

L(3'UTR) nt	769
mRNA RefSeq No	NM 001618.3
Ensembl Gene ID	ENSG00000143799
Ensembl Transcript ID	ENST00000366794
SNP ID	rs8679

In publication

PARP1 rs8679 T>C and the SNP is located within a predicted miRNA-binding site.

In dbSNP

AAAAAGAGCTTTCCTTCTCCAGGAA[C/T]ACTGAACATGGGAGCTCTTGAAATA
NM_001618.3:c.*607T>C

In Ensembl Biomart

NM_001618

>ENSG00000143799|ENST00000366794|PARP1
TTGGGAGAGGTAGCCGAGTCACACCCGGTGGCTCTGGTATGAATTCACCCGAAGCGCTTCTGCACCAACTCACCTGGCCGCTAAGTTGCTGATG
GGTAGTACCTGTACTAAACCACCTCAGAAAGGATTTTACAGAAACGTGTTAAAGGTTTTCTCTAACTTCTCAAGTCCCTTGTTTTGTGTTGTGT
CTGTGGGGAGGGGTTGTTTTGGGGTTGTTTTTGTTTTTTCTTGCCAGGTAGATAAAACTGACATAGAGAAAAGGCTGGAGAGAGATTCTGTTGC
ATAGACTAGTCCTATGGAAAAAACCAAGCTTCGTTAGAATGTCTGCCTTACTGGTTTCCCCAGGGAAGGAAAAATACACTTCCACCCTTTTTTC
TAAGTGTTCGTCTTTAGTTTTGATTTTGGAAAGATGTTAAGCATTTATTTTTAGTTAAAAATAAAAACTAATTTCATACTATTTAGATTTTCTT
TTTTATCTTGCACTTATTGTCCCCTTTTTAGTTTTTTTTGTTTGCCTCTTGTGGTGAGGGGTGTGGGAAGACCAAAGGAAGGAACGCTAACAAT
TTCTCATACTTAGAAACAAAAAGAGCTTTCCTTCTCCAGGAA[C/T]ACTGAACATGGGAGCTCTTGAAATATGTAGTATTAAAAGTTGCATTT
GAAATTCTTGACTTTCTTATGGGCACTTTTGTCTTCCAAATTAAAACTCTACCACAAATATACTTACCCAAGGGCTAATAGTAATACTCGATTA
AAAATGCAGATGCCTTCTCTA

Other available PARP1 3'UTR transcripts in Biomart:

>ENSG00000143799|ENST00000366792 Sequence of length 80
>ENSG00000143799|ENST00000366790|PARP1 Sequence of length 44

Average Sequence length: 302

Unavailable Sequences : ENST00000463968, ENST00000469663, ENST00000491816, ENST00000629232, ENST00000498787, ENST00000490921, ENST00000468608.

WFS1

L(3'UTR) nt	797
mRNA RefSeq No	NM_001145853
Ensembl Gene ID	ENSG00000109501
Ensembl Transcript ID	ENST00000226760
SNP ID	rs1046322

In publication

The rs1046322 SNP generates a mismatch in the target site of the seed region of miR-668.

In dbSNP

TCCACCCTGAGCCTGACCTTTCTGA[A/G]TGACATGGGTGTGCCAGGCTAGACT
NM_001145853.1:c.*253G>A
NM_006005.2:c.*253G>A
NM_006005.3:c.*253G>A

In Ensembl Biomart

NM_001145853
>ENSG00000109501|ENST00000226760|WFS1
GGATGGTCCGCCACGAGGAGCTTCCAGTGCATGTTGCCATGAGGCCTTTCCCCAGTGTGGCCCCAGCCCGACAGGCATGCACCAGTGCCGCCTG
TGCCCACGTGTGCAGACTGTGGCTGCAGAGACCTTGCGACCATGTGTAGATTGCGTGGACCCCGACAAAGGGAAGGCTGCTGTGTAGCTCTGTC
CACTCTGAATACCAAGTGTGTTGGGAATTGCATGCCATCTCCACCCTGAGCCTGACCTTTCTGA[A/G]TGACATGGGTGTGCCAGGCTAGA
CTAGGAGGTTCCGGTGTCTGGAAAAGCACTTTACAGATGAGATTCCCTCTCCTCCCCCACCTTCAAGCACCCTGTTCCCTCTTTCTTTCTTTTG
TGTTGGATTTGTTTAAAAACCAAATAAGCATCTGTGTAACCTCCACAGTAGCATTTCTTATTTGTTTGGTCACTGCTACACCTTAGCAGCTCTT
CCCCTTTCCTGGGGGATGTGCACGGCAGCTTGAGCCTGTCACGTGGTCAAGGCCCGGCCCCATCAGAGGCTGGGGGAGGCGGCACATTGGCAGT
GTGTCACACTGAGCTGGGCACCACAGGCTGCCTCATGACCCTCCTGTCCAGCAGGTAGTGGGTGAATGTGTGAAGGTCTTGCCTGAATCCATCA
GGACTTGGGAAACAGAGAACCCTGTGGGGGCGGCTGTGGGGGAGGTCCCTGCCAGTGTTTAGAAGAGCCTGACTGTGTTCAGTGCCTTGGAGCA
GAAAGCCAGGGTCCTGAGTGGCTGAAATAAAAGCCTCTGGTGGAACCTGCA

NM_006005
>ENSG00000109501|ENST00000226760|WFS1 same sequence

Other available 3'UTR transcripts in Biomart:

>ENSG00000109501|ENST00000503569|WFS1 Sequence of length 414

Average Sequence length: 605

Unavailable Sequences : ENST00000507765, ENST00000506588, ENST00000513395, ENST00000506362.

MSX1

L(3'UTR) nt	790
mRNA RefSeq No	NM_002448
Ensembl Gene ID	ENSG00000163132
Ensembl Transcript ID	ENST00000382723
SNP ID	rs12532

In publication

rs12532 lies within the given potential binding site of miR-3649.

In dbSNP

GCAGAGAGGTTAACAGATTTATCTA[A/G]GGTCCCCAGCAGAATTGACAGTTGA

NM_002448.3:c.*276A>G

In Ensembl Biomart

NM_002448

>ENSG00000163132|ENST00000382723|MSX1

AGGGTCCCAGGTCGCCCACCTGTGGGCCAGCCGATTCCTCCAGCCCTGGTGCTGTACCCCCGACGTGCTCCCCTGCTCGGCACCGCCAGCCGCC
TTCCCTTTAACCCTCACACTGCTCCAGTTTCACCTCTTTGCTCCCTGAGTTCACTCTCCGAAGTCTGATCCCTGCCAAAAAGTGGCTGGAAGAG

TCCCTTAGTACTCTTCTAGCATTTAGATCTACACTCTCGAGTTAAAGATGGGGAAACTGAGGGCAGAGAGGTTAACAGATTTATCTA[A/G]G
GTCCCCAGCAGAATTGACAGTTGAACAGAGCTAGAGGCCATGTCTCCTGCATAGCTTTTCCCTGTCCTGACACCAGGCAAGAAAAGCGCAGAGA
AATCGGTGTCTGACGATTTTGGAAATGAGAACAATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAGAGAAAAAAAAGACT
AGCCAGCCAGGAAGATGAATCCTAGCTTCTTCCATTGGAAAATTTAAGACAAGTTCAACAACAAAACATTTGCTCTGGGGGGCAGGGAAAACAC
AGATGTGTTGCAAAGGTAGGTTGAAGGGACCTCTCTCTTACCAGTACCAGAAACACAATTGTAAAATTAAAAAAAAAAAAAAAACTCTTTCTATT
TAACAGTACATTTGTGTGGCTCTCAAACATCCCTTTGGAAGGGATTGTGTGTACTATGTAATATACTGTATATTTGAAATTTTATTATCATTTA
TATTATAGCTATATTTGTTAAATAAATTAATTTTAAGCTACAA

No other available 3'UTR transcripts in Biomart.

Unavailable Sequences : ENST00000468421

EFNA1

L(3'UTR) nt	843
mRNA RefSeq No	NM_182685
Ensembl Gene ID	ENSG00000169242
Ensembl Transcript ID	ENST00000368406
SNP ID	rs12904

In publication

The rs12904 G > A disrupts the perfect match between the seed region of miR-200c and the binding site in *EFNA1* 3'UTR (the SNP position is 154 in the publication).

In dbSNP

CACCTAGCAGCCTCAAAACGGGTCA[A/G]TATTAAGGTTTTCAACCGGAAGGAG
NM_001122837.1:c.-1732G>A
NM_001122839.1:c.-1691G>A
NM_001287586.1:c.-1732G>A
NM_001287587.1:c.-1732G>A
NM_001287588.1:c.-1732G>A
NM_001287589.1:c.-1732G>A
NM_001287590.1:c.-1451G>A
NM_001287591.1:c.-1394G>A
NM_001287592.1:c.-1366G>A
NM_004428.2:c.*154G>A
NM_018845.3:c.-1691G>A
NM_182685.1:c.*154G>A

In Ensembl Biomart

NM_182685

>ENSG00000169242|ENST00000368406|EFNA1
AGGTGTATGCCACACCTGGCCTTAAAGAGGGACAGGCTGAAGAGAGGGACAGGCACTCCAAACCTGTCTTGGGGCCACTTTCAGAGCCCCCAGC
CCTGGGAACCACTCCCACCACAGGCATAAGCTATCACCTAGCAGCCTCAAAACGGGTCA[A/G]TATTAAGGTTTTCAACCGGAAGGAGGCC
AACCAGCCCGACAGTGCCATCCCCACCTTCACCTCGGAGGGATGGAGAAAGAAGTGGAGACAGTCCTTTCCCACCATTCCTGCCTTTAAGCCAA
AGAAACAAGCTGTGCAGGCATGGTCCCTTAAGGCACAGTGGGAGCTGAGCTGGAAGGGGCCACGTGGATGGGCAAAGCTTGTCAAAGATGCCCC
CTCCAGGAGAGAGCCAGGATGCCCAGATGAACTGACTGAAGGAAAAGCAAGAAACAGTTTCTTGCTTGGAAGCCAGGTACAGGAGAGGCAGCAT
GCTTGGGCTGACCCAGCATCTCCCAGCAAGACCTCATCTGTGGAGCTGCCACAGAGAAGTTTGTAGCCAGGTACTGCATTCTCTCCCATCCTGG
GGCAGCACTCCCCAGAGCTGTGCCAGCAGGGGGGCTGTGCCAACCTGTTCTTAGAGTGTAGCTGTAAGGGCAGTGCCCATGTGTACATTCTGCC
TAGAGTGTAGCCTAAAGGGCAGGGCCCACGTGTATAGTATCTGTATATAAGTTGCTGTGTGTCTGTCCTGATTTCTACAACTGGAGTTTTTTTA
TACAATGTTCTTTGTCTCAAAATAAAGCAATGTGTTTTTTCGGACATGCTTTTCTGCCACTCCATATTAAAACATATGACCATTGAGTCCCTGC
TAA

NM_004428
>ENSG00000169242|ENST00000368407|EFNA1 sequence of length 790

NM_001122837

>ENSG00000169241|ENST00000368401|SLC50A1

NM_001122839

>ENSG00000169241|ENST00000303343|SLC50A1

NM_001287586
>ENSG00000169241|ENST00000622581|SLC50A1

NM_001287587

>ENSG00000169241|ENST00000368404|SLC50A1

NM_001287588
>ENSG00000169241|ENST00000368401|SLC50A1

NM_001287589
NM_001287590
>ENSG00000169241|ENST00000484157|SLC50A1

NM_001287591
NM_001287592
NM_018845
>ENSG00000169241|ENST00000368404|SLC50A1

No other available 3'UTR transcripts in Biomart.

Average Sequence length: 816

Unavailable Sequences : ENST00000474413, ENST00000469878, ENST00000497282

IL-23R

L(3'UTR) nt	851
mRNA RefSeq No	NM_144701.2
Ensembl Gene ID	ENSG00000162594
Ensembl Transcript ID	ENST00000347310
SNP ID	rs10889677

In publication

The IL-23R 3'UTR (RefSeq NM_144701.2) is given. The SNP falls within the seed region of the target site(the binding site is showed in a figure).

In dbSNP

TTTAATTTTAGCCATTCTTCTGCCT[A/C]ATTTCTTAAAATTAGAGAATTAAGG

NM_144701.2:c.*309C>A

In Ensembl Biomart

NM_144701

>ENSG00000162594|ENST00000347310|IL23R
AGCTGTGTGGTCAAAATCAATATGAGAAAGCTGCCTTGCAATCTGAACTTGGGTTTTCCCTGCAATAGAAATTGAATTCTGCCTCTTTTTGAAA
AAAATGTATTCACATACAAATCTTCACATGGACACATGTTTTCATTTCCCTTGGATAAATACCTAGGTAGGGGATTGCTGGGCCATATGATAAG
CATATGTTTCAGTTCTACCAATCTTGTTTCCAGAGTAGTGACATTTCTGTGCTCCTACCATCACCATGTAAGAATTCCCGGGAGCTCCATGCCT
TTTTAATTTTAGCCATTCTTCTGCCT[A/C]ATTTCTTAAAATTAGAGAATTAAGGTCCCGAAGGTGGAACATGCTTCATGGTCACACATAC
AGGCACAAAAACAGCATTATGTGGACGCCTCATGTATTTTTTATAGAGTCAACTATTTCCTCTTTATTTTCCCTCATTGAAAGATGCAAAACAG
CTCTCTATTGTGTACAGAAAGGGTAAATAATGCAAAATACCTGGTAGTAAAATAAATGCTGAAAATTTTCCTTTAAAATAGAATCATTAGGCCA
GGCGTGGTGGCTCATGCTTGTAATCCCAGCACTTTGGTAGGCTGAGGTAGGTGGATCACCTGAGGTCAGGAGTTCGAGTCCAGCCTGGCCAATA
TGCTGAAACCCTGTCTCTACTAAAATTACAAAAATTAGCCGGCCATGGTGGCAGGTGCTTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAG
AATCACTTGAACCAGGAAGGCAGAGGTTGCACTGAGCTGAGATTGTGCCACTGCACTCCAGCCTGGGCAACAAGAGCAAAACTCTGTCTGGAAA
AAAAAAAAAA

Other available 3'UTR transcripts in Biomart:

>ENSG00000162594|ENST00000425614|IL23R Sequence of length 120
>ENSG00000162594|ENST00000395227|IL23 Sequence of length 543
>ENSG00000162594|ENST00000473881|IL23R Sequence of length 1263
>ENSG00000162594|ENST00000637002|IL23R Sequence of length 39

Average Sequence length: 563

RYR3

L(3'UTR) nt	880
mRNA RefSeq No	NM_001036.3
Ensembl Gene ID	ENSG00000198838
Ensembl Transcript ID	ENST00000634891
SNP ID	rs1044129

In publication

The SNP (rs1044129 A→G) is located near the miRNA-367 binding site (the binding region is showed in a figure).

In dbSNP

GTTTAGGTGAATCTCCTCAAATACA[A/G]TGAAGTGCCCACTGCAATAAAGTAA

NM_001036.4:c.*839A>G
NM_001243996.2:c.*839A>G
NM_020371.2:c.*386+162T>C
NM_020371.2:c.*548T>C

In Ensembl Biomart

NM_001036

>ENSG00000198838|ENST00000634891|RYR3
ATCTGAATCAAAGAAGCGCGACAATTCTGGACAGTCAACTTCCCATGAAATAAAGTCCCCTTTTTACAGTTCTGCAACATATCTGAAATGTGAC
ATTTTCTAAATGCCTCCCTTAAAAAAAAAACTGCTGAAAATCTGTGCTATTTTGAAATTGATTTGGCTTTTTGTGCCTAATGGACATACACTGT
GGGAGAGAACCTGTCAAAATGTCGAAGAAGGAAGGCGAAGAATCAAGTAATCTCTAGGCAAATGCCTTCAAGTTTTCCAGTTCTGAGGTAACTA
GTTCAGTTTGTTGGGATGGAAGCATGAAGGAAAGGGCTAGAGAAGTATGAAATCTCGAATGTGTAATACCTGAAAATTTAAACACTTGAATGTC
ATCATGGTATCCAACTTGTGACTCATAGGGTCTGAACTCACTCCAAAAGATAATAACTGCAGTCTAATTTTTCCCATGGTACTTGCTAGTGACT
GTATCCAGAAAAGCTTTAAGCAGTTAAAGAAACAGAAAAAAACCGACACTTTGTCGACACTGAAATATCGATTAAGTGCCTTAAAACCTCTTTA
GACATAGCTATGCAAGTTTTTTATGTTTGTGTTCCAGAAGGACAGTTCCATTCATTAGTTGTGATCTTCCGTCTTACTTTATGAAACTGCACTT
GAAGGTTATTCATACAAGTTTTTTTAGTAACAGCTGTCAGTCAACTGCTGTTATTAGAAGAAAAGTACTGTACTGAAAATTCAGAAAAAAAATC
TCAACCTTATGCCAAAATGGAGTAATGCTTTATGGTCCCTTGTAAGTAGTGGAGCTGCTCTGTTTAGGTGAATCTCCTCAAATACA[A/G]T
GAAGTGCCCACTGCAATAAAGTAATACGTACCAATAAAAA

NM_001243996

>ENSG00000198838|ENST00000415757|RYR3 Same sequence of length 875

NM_020371

>ENSG00000169857|ENST00000306730|AVEN

Other available 3'UTR transcripts in Biomart:

>ENSG00000198838|ENST00000389232|RYR3 Sequence of length 872
>ENSG00000198838|ENST00000622037|RYR3 Sequence of length 872
>ENSG00000198838|ENST00000637948|RYR3 Sequence of length 2173
>ENSG00000198838|ENST00000634418|RYR3 Sequence of length 498
>ENSG00000198838|ENST00000634730|RYR3 Sequence of length 895
>ENSG00000198838|ENST00000635790|RYR3 Sequence of length 5224
>ENSG00000198838|ENST00000636845|RYR3 Sequence of length 1610
>ENSG00000198838|ENST00000636568|RYR3 Sequence of length 851

Average Sequence length: 1475

Unavailable Sequences : ENST00000560791, ENST00000557931, ENST00000638052, ENST00000636656, ENST00000635749, ENST00000638038, ENST00000637522, ENST00000638145, ENST00000638032, ENST00000636497, ENST00000637072, ENST00000559333, ENST00000559917, ENST00000558060, ENST00000635875, ENST00000637201, ENST00000638085, ENST00000636753, ENST00000634750, ENST00000635842, ENST00000636878, ENST00000636417, ENST00000637615, ENST00000636583, ENST00000637984.

AGTR1

L(3'UTR) nt	888
mRNA RefSeq No	NM_009585
Ensembl Gene ID	ENSG00000144891
Ensembl Transcript ID	ENST00000497524
SNP ID	rs5186

In publication

The rs5186 (1166A/C), is located in the region that interacts with the miR-155 seed (the binding region is showed in a figure where the SNP disrupts a perfect seed match).

In dbSNP

TGCAGCACTTCACTACCAAATGAGC[A/C]TTAGCTACTTTTCAGAATTGAAGGA
NM_000685.4:c.*86A>C
NM_004835.4:c.*86A>C
NM_009585.3:c.*86A>C
NM_031850.3:c.*86A>C
NM_032049.3:c.*86A>C

In Ensembl Biomart

NM_009585

>ENSG00000144891|ENST00000497524|AGTR1 Sequence of length 888

CATGTTCGAAACCTGTCCATAAAGTAATTTTGTGAAAGAAGGAGCAAGAGAACATTCCTCTGCAGCACTTCACTACCAAATGAGC[A/C]TT
AGCTACTTTTCAGAATTGAAGGAGAAAATGCATTATGTGGACTGAACCGACTTTTCTAAAGCTCTGAACAAAAGCTTTTCTTTCCTTTTGCAAC
AAGACAAAGCAAAGCCACATTTTGCATTAGACAGATGACGGCTGCTCGAAGAACAATGTCAGAAACTCGATGAATGTGTTGATTTGAGAAATTT
TACTGACAGAAATGCAATCTCCCTAGCCTGCTTTTGTCCTGTTATTTTTTATTTCCACATAAAGGTATTTAGAATATATTAAATCGTTAGAGGA
GCAACAGGAGATGAGAGTTCCAGATTGTTCTGTCCAGTTTCCAAAGGGCAGTAAAGTTTTCGTGCCGGTTTTCAGCTATTAGCAACTGTGCTAC
ACTTGCACCTGGTACTGCACATTTTGTACAAAGATATGCTAAGCAGTAGTCGTCAAGTTGCAGATCTTTTTGTGAAATTCAACCTGTGTCTTAT
AGGTTTCCACTGCCAAAACAATGCCCGTAAGATGGCTTATTTGTATAATGGTGTTACTAAAGTCACATATAAAAGTTAAACTACTTGTAAAGGT
GCTGCACTGGTCCCAAGTAGTAGTGTCTTCCTAGTATATTAGTTTGATTTAATATCTGAGAAGTGTATATAGTTTGTGGTAAAAAGATTATATA
TCATAAAGTATGCCTTCCTGTTTAAAAAAAGTATATATTCTACACATATATGTATATGTATATCTATATCTCTAAACTGCTGTTAATTGATTAA
AATCTGGCAAAGTTATATTTACTTTAAAATAAAATAATTTTATTGCAA

NM_000685

>ENSG00000144891|ENST00000349243|AGTR1 Same sequence of length 888

NM_004835

>ENSG00000144891|ENST00000418473|AGTR1 Same sequence of length 886

NM_031850

>ENSG00000144891|ENST00000418473|AGTR1 Same sequence of length 886

NM_032049

>ENSG00000144891|ENST00000402260|AGTR1 Same sequence of length 886

Other available 3'UTR transcripts in Biomart:

>ENSG00000144891|ENST00000404754|AGTR1 sequence of length 893
>ENSG00000144891|ENST00000461609|AGTR1 sequence of length 132
>ENSG00000144891|ENST00000474935|AGTR1 sequence of length 132
>ENSG00000144891|ENST00000475347|AGTR1 sequence of length 132

Average Sequence length: 636

Unavailable Sequences : ENST00000475166

FGF20

L(3'UTR) nt	903
mRNA RefSeq No	NM_019851.2
Ensembl Gene ID	ENSG00000078579
Ensembl Transcript ID	ENST00000180166
SNP ID	rs12720208

In publication

The SNP rs12720208 lies within a predicted binding site for miR-433. The binding region is showed in a figure where the allele C base paired with G in Watson-Crick mode whereas allele T wobble base paired with G.

In dbSNP

We selected for rs12720208 the dbSNP submission ss20399075 instead of ss28476621, which results in a complete match to the 3'UTR NM_019851.2:c.*182C>T (ENST00000180166).

AAAAAAATCTTGACTAGAAATAGAT[C/T]ATGATCACTCTTTATATGTGGATTA

NM_019851.2:c.*182C>T

In Ensembl Biomart

NM_019851

>ENSG00000078579|ENST00000180166|FGF20
AGTGCGATAGTGACATTATGGAAGAGTCAAACCACAACCATTCTTTCTTGTCATAGTTCCCATCATAAAATAATGACCCAAGGAGACGTTCAAA
ATATTAAAGTCTATTTTCTACTGAGAGACTGGATTTGGAAAGAATATTGAGAAAAAAAACCAAAAAAAATTTTGACTAGAAATAGAT[C/T]
ATGATCACTCTTTATATGTGGATTAAGTTCCCTTAGATACATTGGATTAGTCCTTACCAGTAGACTGAAGCCAAAAATCTGTTAAACTGTGGAT
AATGGGAACCCTCACATGGTTGCTGCTGGTGCCTCACCTCTCTGGATTATGTTTACTTTAAAATTTATGTTAAAAATAGAGACTTCATGTTGAA
GAAATGGATTGACTTAGCCAAGAGCATTGCTACATAAATTCTGAGGACCTTGTAGAATTCACAGGTTATGGCATTATATGTTTTAAGAAAAAAA
AAAACTAAACAACACCTGGGTGAAACATGAACTATGAAACATGCCACACCAAGACAGAACACTTTTTAAAAGAATGGACCTATTTATACACTTT
CAATTTGAAAATATTTATTATATATATTTATTTATTAAAGAGTTTATTTTTTACTGGGATATGAAGATAATACAAAGAGTAAAAAACAACAAAA
CTTCCGTTATTGTGCCATTACTTTATTGTGGTGTCTTGCTCTGTAAAGAGGACATTATTATACCACCTTAACCATAGAATCAGATTTTGAATTT
CTTTTTAAAATATAGTATAGATTATATTTTTTATGCAGTCAATTGCCTTCTAAATGTAACGTTGGTTTCTTTTGGCCTAGAAAAATGCCTCATT
GATTTGTATTAAATTCAACACACATTCAAAAGGGAATTAAACATTTTCATGCAGTGTTTTGGC

Other available 3'UTR transcripts in Biomart:

>ENSG00000078579|ENST00000519941|FGF20 sequence of length 94

Average Sequence length: 498

DROSHA

L(3'UTR) nt	935
mRNA RefSeq No	NM_013235
Ensembl Gene ID	ENSG00000113360
Ensembl Transcript ID	ENST00000511367
SNP ID	rs10719

In publication

The RefSeq ID of the used transcript for DROSHA is given (NM_013235). The SNP rs10719TC was adjacent to the hsa-miR-27b binding site in DROSHA 3'UTR (binding region given in a figure).

In dbSNP

GCCTAGTTTTCCTGCAGACAATGAA[C/T]GAAGTGTGCTCATTGAAATAAAATA

NM_001100412.1:c.*92T>C

NM_013235.4:c.*92T>C

In Ensembl Biomart

NM_001100412

>ENSG00000113360|ENST00000513349|DROSHA Sequence of length 298

NM_013235

>ENSG00000113360|ENST00000511367|DROSHA
AGGAGGGCATGCAAGTGTGGAGTATTTACTTGCTCAGTAACTGTGACTGTTGTCTATTGAGACCTAGCCTAGTTTTCCTGCAGACAATGAA[C/
T]GAAGTGTGCTCATTGAAATAAAATACAGAGTCAAATCGCTATTGTTGTTTTAATGATCTGTTTTTAGCTGGATGGTCTTTATTACAAAGTAT

TAGATTTTTCTTCTATTTAACGGAAAACTTGACTTTGGTGAATGTGCATTACTTCCTTTTATTTTGCTCTTTAAATAATAAAATTCAAGAAGCA
TATTCTATGTGGAATAGATCCTGTTTTTCCATCTGTGTCCCAGATTGTGACCCTAGACTTTCAATTGACAAGTAAAAAATTGACTTTACTAGAC
ATTTTGACTGTGCTCTAGTAACATCTATCCTTTTTCAAATCTCTGGATTTTTAAGTAGATTGTTCAGCTTTCATCCGGTGGCTGTTCATCAAGT
TATCAGCTGCAAATATTGAACTTACCTCTCTCTAAGCAGTGAGTGTTTTGTAGAAGGAATCCGTTTAACAATTAATTGGCTAATGGGAGAAGGG
GAAAGACTGATATTCAAGTCATACAGATTCTTTGAATCATTAGAATAGGAGAGAAATCATGATTCTAAGCCAGGCCACACTTTAAACCAAGTGC
TCTCACCCTGGGGTTAGTGGAACCTTTAAGAAGTTAATGAACAGACTTCAAGGAAGTCAAAAACCTCCCAATACTATATTCAATTTTCTGTGTG
TGTTTGAGATTTGAGAGGGCATGGGGAGCAGGAAGGAGGAGGGTTTATAGCTTTTATCAGCCTCTCTAAGTGGGCCCTGCAGTAAAAGGCTAAC
ATGACATTCAAAGACATAACATTTTAAAAAAAGTTATTCCAAACTAAACATCACTGGTTTCTTATTAATAAAGGCAAAACTTCTTTGTAAAAC

Other available 3'UTR transcripts in Biomart:

>ENSG00000113360|ENST00000442743|DROSHA sequence of length 930
>ENSG00000113360|ENST00000344624|DROSHA sequence of length 930

Average Sequence length: 773

Unavailable Sequences : ENST00000509608, ENST00000507438, ENST00000510178, ENST00000504133, ENST00000509067, ENST00000512885, ENST00000504361, ENST00000510375, ENST00000512124, ENST00000512166, ENST00000514927, ENST00000507174, ENST00000512076, ENST00000511778, ENST00000505601, ENST00000512302, ENST00000511803.

HOXB5

L(3'UTR) nt	952
mRNA RefSeq No	NM_002147.3
Ensembl Gene ID	ENSG00000120075
Ensembl Transcript ID	ENST00000239151
SNP ID	rs9299

In publication

The used RefSeq ID of the HOXB5 3'UTR is given (NM_002147.3). The binding region is given in a figure, showing the SNP location in the binding site next to the seed region.

In dbSNP

GCCTTGTGTTACGATATTTCGTTTG[A/G]TCTTAGGTCTTCCTGTGGCTCCCTC
NM_002147.3:c.*141G>A

In Ensembl Biomart

NM_002147

>ENSG00000120075|ENST00000239151|HOXB5
GCCCGCCCAGAGGAGCCCAGCGGCCCAAGAGCCCGTGCCACCCCCAGCCCTGGCCCCTCCAATCCTCCCCGCTCTGCCGCCGCCCGCTGGGGAC
CGGTTCCCACAAGCCTGCCTCGCCTTGTGTTACGATATTTCGTTTG[A/G]TCTTAGGTCTTCCTGTGGCTCCCTCTCTCCTGGACTGGTTATC
TTGTTATTATTGTTAATAATAATTATTATTATTATTTTCCTTCCATGCTCCCAACTCCCTTCTGCTTGTCCCAAATCCGCCAGTGTTTCTGAAT
GTTTGTGTCTGTGGTTGCAGTCTTTCCCCCAGGAAAAAAAAAAAAAGAAATTCGCATGTTTAATGTGAACTCTCCCCTCCCCATCTGTGTTCTA
ACTTATTTATAAAAAGATGATCGCTGTATTTTGAGTTTCAGCTGGAAACTTCTGTAAGGGGCAGCAGTTGAGGTGGGGTAGTGCCGCAGTGGGG
TCAAGCTGAGCTGGCTTCGGAGATGGAGTCCCTTTTCATTCTCCTCCTCCCTCCTCACTCCCTAGGCCCAAGTCTCCTAGGGGCTTGGTCC
TAGGGTGGGAAGGGGCTAGGGAGGACCAAAGGGATGGTATTGAGAAGAGAGAAAGAAGATAGTGAGATTTAAGTTCCTGCTGCCTGGGTAGGCC
CCACAAGGCCTGGTCTGGGAGTATACGGAAACAAAAATGATCCTCAGTGCAAAATGTCTTGTGTATTTCTCTGTGAATCCATGGGTCTGGCTAG
AGGGCCCAAAGCTTGTAAATATGGGGATAGTCTGGGTCAGACCCATCTCTCCCTTACCCATCTTGCTTCCAAGACCATTTGTAGTGAGCGAGTG
GATGCTGTGCTACGTGTGAAATCTGTCTTTGCGGGGCCTGTCTCAGTGATTCGCTTTTGGTATTTGTTTGTAGCTTTCCTGGAAGTCAAATAAA
TGTTTCCCCCACTCCA

No other available 3'UTR transcripts in Biomart.

RAD51

L(3'UTR) nt	978
mRNA RefSeq No	NM 002875.4

Ensembl Gene ID ENSG00000051180
Ensembl Transcript ID ENST00000267868
SNP ID rs7180135

In publication

RAD51 rs7180135 A>G and the SNP is located within a predicted miRNA-binding site.

In dbSNP

GAGTTTTAAGTCCAGCTTGGCCAAG[A/G]TGGTGAAATCCCATCTCTACAAAAA
NM_001164269.1:c.*718G>A
NM_001164270.1:c.*773G>A
NM_002875.4:c.*718G>A
NM_133487.3:c.*718G>A

In Ensembl Biomart

NM_001164269

>ENSG00000051180|ENST00000382643|RAD51 Sequence of length 415

NM_001164270

>ENSG00000051180|ENST00000423169|RAD51 Sequence of length 469

NM_002875

>ENSG00000051180|ENST00000267868|RAD51
ATCATTGGGTTTTTCCTCTGTTAAAAACCTTAAGTGCTGCAGCCTAATGAGAGTGCACTGCTCCCTGGGGTTCTCTACAGGCCTCTTCCTGTTG
TGACTGCCAGGATAAAGCTTCCGGGAAAACAGCTATTATATCAGCTTTTCTGATGGTATAAACAGGAGACAGGTCAGTAGTCACAAACTGATCT
AAAATGTTTATTCCTTCTGTAGTGTATTAATCTCTGTGTGTTTTCTTTGGTTTTGGAGGAGGGGTATGAAGTATCTTTGACATGGTGCCTTAGG
AATGACTTGGGTTTAACAAGCTGTCTACTGGACAATCTTATGTTTCCAAGAGAACTAAAGCTGGAGAGACCTGACCCTTCTCTCACTTCTAAAT
TAATGGTAAAATAAAATGCCTCAGCTATGTAGCAAAGGGAATGGGTCTGCACAGATTCTTTTTTTCTGTCAGTAAAACTCTCAAGCAGGTTTTT
AAGTTGTCTGTCTGAATGATCTTGTGTAAGGTTTTGGTTATGGAGTCTTGTGCCAAACCTACTAGGCCATTAGCCCTTCACCATCTACCTGCTT
GGTCTTTCATTGCTAAGACTAACTCAAGATAATCCTAGAGTCTTAAAGCATTTCAGGCCAGTGTGGTGTCTTGCGCCTGTACTCCCAGCACTTT
GGGAGGCCGAGGCAGGTGGATCGCTTGAGCCCAGGAGTTTTAAGTCCAGCTTGGCCAAG[A/G]TGGTGAAATCCCATCTCTACAAAAAATGCA
GAACTTAATCTGGACACACTGTTACACGTGCCTGTAGTCCCAGCTACTCGATAGCCTGAGGTGGGAGAATCACTTAAGCCTGGAAGGTGGAAGT
TGCAGTGAGTCGAGATTGCACTGCTGCATTCCAGCCAGGGTGACAGAGTGAGACCATGTTTCAAACAAGAAACATTTCAGAGGGTAAGTAAACA
GATTTGATTGTGAGGCTTCTAATAAAGTAGTTATTAGTAGTGAATGTGCTGTTTATAGCAATTATTGCAGTGCAAGCTATTTCAAGACAGGGTT
TCCATAATCTTTTTGGCACTGTATAGGGGTGATCAGTTTCTGTTGCTTCAAGATTTGAAACCAGAAAGGAAAGTCCCACTTGCAGATGATTGTG
CTTAAAAGCTAATGGAAAAATAAAAGAAAGACATACC

NM_133487

>ENSG00000051180|ENST00000532743|RAD51|1515 Sequence of length 238

Other available 3'UTR transcripts in Biomart:

>ENSG00000051180|ENST00000531277|RAD51 353
>ENSG00000051180|ENST00000525066|RAD51 848
>ENSG00000051180|ENST00000533741|RAD51 228
>ENSG00000051180|ENST00000557850|RAD51 412

Average Sequence length: 493

Unavailable Sequences : ENST00000526763 , ENST00000527860

REV3L

L(3'UTR) nt 985
mRNA RefSeq No NM_002912
Ensembl Gene ID ENSG00000009413
Ensembl Transcript ID ENST00000358835

SNP ID rs465646

In publication

The *REV3L* 3'UTR and 460 T>C SNP (rs465646) is located within the predicted upstream miRNA-binding site.

In dbSNP

TACACTTTTTTACTTTTCAAACGAG[T/C]AAAATAATGTGCAATGATTTTTATA
NM_001286431.1:c.*461C>T
NM_001286432.1:c.*461C>T
NM_002912.4:c.*461C>T

In Ensembl Biomart

NM_002912

>ENSG00000009413|ENST00000358835|REV3L
ATTGTCAATATCACAGTATTACAGGTGCTATTTTTTTCAGTGCTTACCACTAAACTGTTGTGCATGGTGCTTTTTAACTTTCATCGAGTCAAGGATGTTCACTGTCT
GTTATCTGAAGACTATGAAGACTTCTATGCTAACCGAATTAAAATGTACTTGTTGATCTCTGAATAGCTCACTTCTTACAATGTACAAATTCCTCATTCTGTCACCT
TTTAAACATTGTTTTATAATGCAGGTGTTGGATTTGCTCCAGTATGTGTACCATCTTGTAAATTCATTTGAGTAGATCATGTTTACTTCCCAGTGGAAGGAGCACTG
AAAACCTCTTAAAGAAAAAGCATTTGTGTGTTTTCCTTGAACTGTCTGTATCAAGACGTGTTACTTCGAGATATCCATTCACTTTATAATTTTGACTGCAAAATATT
TTGTAAATACACTTTTTTACTTTTCAAACGAG[T/C]AAAATAATGTGCAATGATTTTTATACAAATGATTTTCAAGTTGTTTGGTATATTTCCTCTAGGTTTTGCT
TGACTCAAAGTAGATCGTTATTTTGATCAAACTGTGCAAACAGTAGTACCACGTGTAGCATTTTGAAACATTATTTTTTTTTAAAAAATGCTGTCTTGCTTTAGCTA
TTAATGGGGCATTGTGAGGAACTGTGCAAAGACATTTTTGTTACAAACCTGTGGGCCTGTTGCAATACTTTAAAAATAAAAAATTTTATTCCATTTGCTTGTTTTGT
ATAGACATTTCTATTGCTTCTAAATATACTTAAAATATTTTCTTTCCTTATGTACTGTACAGTTAATCTTATTTGCCATCATCTTGAACACAAAATGTGTATTTAGA
ATATTTGTATAACTGTGTAAAATAAAAAAGGAATTATGTGGTCAGTGCATTGTTTTTTAAACTGGAAATCATTTTGTTTTAAAAGTTAATAATGGAAACCATATTAA
AATTGAATAAAATATAAAATAATATA

NM_001286431
>ENSG00000009413|ENST00000435970|REV3L Similar sequence of length 964

NM_001286432
>ENSG00000009413|ENST00000435970|REV3L Similar sequence of length 964

Other available 3'UTR transcripts in Biomart:

>ENSG00000009413|ENST00000413831|REV3L Sequence of length 336
>ENSG00000009413|ENST00000434009|REV3L Sequence of length 10469
>ENSG00000009413|ENST00000422377|REV3L Sequence of length 10362
>ENSG00000009413|ENST00000368802|REV3L Sequence of length 985
>ENSG00000009413|ENST00000368805|REV3L Sequence of length 985

Average Sequence length: 3256

Unavailable Sequences : ENST00000494858, ENST00000492520, ENST00000470871, ENST00000467500, ENST00000462119, ENST00000460981, ENST00000619481

ORAI1

L(3'UTR) nt	1034
mRNA RefSeq No	NM_032790.3
Ensembl Gene ID	ENSG00000276045
Ensembl Transcript ID	ENST00000617316
SNP ID	rs76753792

In publication

ORAI1 transcript (NM_032790) is given and the SNP 14988C/T (rs76753792).

In dbSNP

CTGCCCCAGCCTCACGGACAGCCTG[C/T]GCAGGGGGCTGGGCTTCAGCAAGGG
NM_032790.3:c.*86C>T

In Ensembl Biomart

NM_032790

>ENSG00000276045|ENST00000617316|ORAI1
GCCCATGTGGTCTGGGCCCTTCCAGTGCTTTGGCCTTACGCCCTTCCCCTTGACCTTGTCCTGCCCCAGCCTCACGGACAGCCTG[C/T]GCAG
GGGGCTGGGCTTCAGCAAGGGGCAGAGCATGGAGGGAAGAGGATTTTTATAAGAGAAATTTCTGCACTTTGAAACTGTCCTCTAAGAGAATAAG
CATTTCCTGTTCTTCCAGCTCCAGGTCCACCTCCTGTTGGGAGGCGGTGGGGGGCCAAAGTGGGGCCACACACTCGCTGTGTCCCCTCTCCTCC
CCTGTGCCAGTGCCACCTGGGTGCCTCCTCCTGTCCTGTCCGTCTCAACCTCCCTCCCGTCCAGCATTGAGTGTGTACATGTGTGTGTGACACA
TAAATATACTCATAAGGACACCTCCTTCCCGTGTCTTGTATTTGTTGGGCCTGGGCTACTGCTCACCCTGGTTAGGTGAGCCTCTAGGAAAACT
TAAAACAAATTTTAAGCCAGGTATGGTGGCACATACCTGTGGTCTCAGCTATTCAGGAGGCCAAGGCAGGAGGATCTCTTGAGCCCAGGAGTTT
GAGACCCCATCTCAAACAAAAAATACAAAAATTAGCCAGCCACGGCGCCTGCACTTCCAGCTCCTTTGAGAGACTGAGGCAGGAAGATTGCCTA
AGCCCAGGAGGCCAAGTCTGCAGTGAGCTATGGTAACACCACTGCACTCCAACCTGGGCAACAGAGGGAGACTCTGTCTCTAAAAAAATAGAAA
AATTTGCCCTGCATGGTGGCTCACGCCTGTAATCCTAGCCCTTTGGAAGGCCAAGGCGGGCAGATCACTTGAGGTCGGGAGTTCGAGACCAGCC
TGACCAACATGGAGAAACCCCATCTGTACTAAAAATACAAAATTAGCTGGGTTTGGTGGCGCATGCTTGTAATCCCAGCTACTCGGGAGGCTGA
GGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATCGCGCCATTGCACTCCAGCCTGGGCAACAACAGTGAAACTCCG
TCTC

Other available 3'UTR transcripts in Biomart:

>ENSG00000276045|ENST00000611718|ORAI1 Sequence of length 461

Average Sequence length: 747

RAP1A

L(3'UTR) nt	1078
mRNA RefSeq No	NM_002884.2
Ensembl Gene ID	ENSG00000116473
Ensembl Transcript ID	ENST00000369709
SNP ID	rs6573

In publication

SNP rs6573 lays within a given binding site of RAP1A for miR-196a (binding region TGATCTTTTTATCATGATCCTMCCTA). the A allele matches the predicted seed region of miR196a, whereas the C allele represents a C:U mismatch base pairing (showed in a figure).

In dbSNP

ATATTGATCTTTTTATCATGATCCT[A/C]CCTATCAAGCACTAAAAAGTTGAAC
NM_001010935.2:c.*366C>A
NM_001291896.1:c.*366C>A
NM_002884.3:c.*366C>A

In Ensembl Biomart

NM_001010935

>ENSG00000116473|ENST00000369709|RAP1A Sequence of length 4290

NM_001291896

>ENSG00000116473|ENST00000369709|RAP1A Sequence of length 4290

NM_002884

>ENSG00000116473|ENST00000436150 Sequence of length 712

Other available 3'UTR transcripts in Biomart:

>ENSG00000116473|ENST00000356415|RAP1A Sequence of length 712

Average Sequence length: 2216

Unavailable Sequences : ENST00000494982, ENST00000433097

APP

L(3'UTR) nt	1120
mRNA RefSeq No	NM_000484
Ensembl Gene ID	ENSG00000142192
Ensembl Transcript ID	ENST00000346798
SNP ID	T171C

In publication

The SNP region is given in the article (GAATTAATCCACACA) with the binding site. The T171C SNP disrupts the perfect match between the seed region of miR-147and the given binding site (showed in a figure).

Length hAPP 3'UTR ~ 1100 bp in figure 1

In dbSNP

Rs ID unknown.

In Ensembl Biomart

>ENSG00000142192|ENST00000346798|APP
ACCCCCGCCACAGCAGCCTCTGAAGTTGGACAGCAAAACCATTGCTTCACTACCCATCGGTGTCCATTTATAGAATAATGTGGGAAGAAACAAA
CCCGTTTTATGATTTACTCATTATCGCCTTTTGACAGCTGTGCTGTAACACAAGTAGATGCCTGAACTTGAATTAA[T/A]CCACACATCAGTA
ATGTATTCTATCTCTCTTTACATTTTGGTCTCTATACTACATTATTAATGGGTTTTGTGTACTGTAAAGAATTTAGCTGTATCAAACTAGTGCA
TGAATAGATTCTCTCCTGATTATTTATCACATAGCCCCTTAGCCAGTTGTATATTATTCTTGTGGTTTGTGACCCAATTAAGTCCTACTTTACA
TATGCTTTAAGAATCGATGGGGGATGCTTCATGTGAACGTGGGAGTTCAGCTGCTTCTCTTGCCTAAGTATTCCTTTCCTGATCACTATGCATT
TTAAAGTTAAACATTTTTAAGTATTTCAGATGCTTTAGAGAGATTTTTTTTCCATGACTGCATTTTACTGTACAGATTGCTGCTTCTGCTATAT
TTGTGATATAGGAATTAAGAGGATACACACGTTTGTTTCTTCGTGCCTGTTTTATGTGCACACATTAGGCATTGAGACTTCAAGCTTTTCTTTT
TTTGTCCACGTATCTTTGGGTCTTTGATAAAGAAAAGAATCCCTGTTCATTGTAAGCACTTTTACGGGGCGGGTGGGGAGGGGTGCTCTGCTGG
TCTTCAATTACCAAGAATTCTCCAAAACAATTTTCTGCAGGATGATTGTACAGAATCATTGCTTATGACATGATCGCTTTCTACACTGTATTAC
ATAAATAAATTAAATAAAATAACCCCGGGCAAGACTTTTCTTTGAAGGATGACTACAGACATTAAATAATCGAAGTAATTTTGGGTGGGGAGAA
GAGGCAGATTCAATTTTCTTTAACCAGTCTGAAGTTTCATTTATGATACAAAAGAAGATGAAAATGGAAGTGGCAATATAAGGGGATGAGGAAG
GCATGCCTGGACAAACCCTTCTTTTAAGATGTGTCTTCAATTTGTATAAATGGTGTTTTCATGTAAATAAATACATTCTTGGAGGAGCA

Other available 3'UTR transcripts in Biomart:

>ENSG00000142192|ENST00000440126|APP	Sequence of length 339
>ENSG00000142192|ENST00000359726|APP	Sequence of length 1119
>ENSG00000142192|ENST00000358918|APP	Sequence of length 60
>ENSG00000142192|ENST00000448850|APP	Sequence of length 390
>ENSG00000142192|ENST00000439274|APP	Sequence of length 338
>ENSG00000142192|ENST00000357903|APP	Sequence of length 1120
>ENSG00000142192|ENST00000354192|APP	Sequence of length 1120
>ENSG00000142192|ENST00000348990|APP	Sequence of length 1120

Average Sequence length: 747

Unavailable Sequences : ENST00000466453, ENST00000548570, ENST00000463070, ENST00000491395, ENST00000462267, ENST00000474136, ENST00000464867, ENST00000415997

IL1A

L(3'UTR) nt	1152
mRNA RefSeq No	NM_000575

Ensembl Gene ID	ENSG00000115008
Ensembl Transcript ID	ENST00000263339
SNP ID	rs3783553

In publication

The rs3783553 SNP lies within a given binding site for human miR-122. rs3783553 occurs in the 7 bp seed sequence of complementarity at has-miR-122 5′ end (alignment showed in a figure).

In dbSNP

AACTTACCTGGGCATTCTTGTTTCA[-/TTCA]ATTCCACCTGCAATCAAGTCCTACA
NM_000575.4:c.*928_*929insTTCA

In Ensembl Biomart

>ENSG00000115008|ENST00000263339|IL1A
GTCTGGAGTCTCACTTGTCTCACTTGTGCAGTGTTGACAGTTCATATGTACCATGTACATGAAGAAGCTAAATCCTTTACTGTTAGTCATTTGC
TGAGCATGTACTGAGCCTTGTAATTCTAAATGAATGTTTACACTCTTTGTAAGAGTGGAACCAACACTAACATATAATGTTGTTATTTAAAGAA
CACCCTATATTTTGCATAGTACCAATCATTTTAATTATTATTCTTCATAACAATTTTAGGAGGACCAGAGCTACTGACTATGGCTACCAAAAAG
ACTCTACCCATATTACAGATGGGCAAATTAAGGCATAAGAAAACTAAGAAATATGCACAATAGCAGTTGAAACAAGAAGCCACAGACCTAGGAT
TTCATGATTTCATTTCAACTGTTTGCCTTCTACTTTTAAGTTGCTGATGAACTCTTAATCAAATAGCATAAGTTTCTGGGACCTCAGTTTTATC
ATTTTCAAAATGGAGGGAATAATACCTAAGCCTTCCTGCCGCAACAGTTTTTTATGCTAATCAGGGAGGTCATTTTGGTAAAATACTTCTTGAA
GCCGAGCCTCAAGATGAAGGCAAAGCACGAAATGTTATTTTTTAATTATTATTTATATATGTATTTATAAATATATTTAAGATAATTATAATAT
ACTATATTTATGGGAACCCCTTCATCCTCTGAGTGTGACCAGGCATCCTCCACAATAGCAGACAGTGTTTTCTGGGATAAGTAAGTTTGATTTC
ATTAATACAGGGCATTTTGGTCCAAGTTGTGCTTATCCCATAGCCAGGAAACTCTGCATTCTAGTACTTGGGAGACCTGTAATCATATAATAAA
TGTACATTAATTACCTTGAGCCAGTAATTGGTCCGATCTTTGACTCTTTTGCCATTAAACTTACCTGGGCATTCTTGTTTCA[-
/TTCA]ATTCCACCTGCAATCAAGTCCTACAAGCTAAAATTAGATGAACTCAACTTTGACAACCATGAGACCACTGTTATCAAAACTTTCTTTT
CTGGAATGTAATCAATGTTTCTTCTAGGTTCTAAAAATTGTGATCAGACCATAATGTTACATTATTATCAACAATAGTGATTGATAGAGTGTTA
TCAGTCATAACTAAATAAAGCTTGCAACAAAATTCTCTGACA

No other available 3′UTR transcripts in Biomart.

CD133

L(3'UTR) nt	1167
mRNA RefSeq No	NM_0011458
Ensembl Gene ID	ENSG00000007062
Ensembl Transcript ID	ENST00000540805
SNP ID	rs2240688

In publication

CD133 rs2240688A to rs2240688C transition gained a new binding of the microRNA hsa-miR-135b.

In dbSNP

CCACTTGGAAATGGCATGCAAAAGC[A/C]ATCATAGAGAAACCTGCGTAACTCC
NM_001145847.1:c.*667A>C
NM_001145848.1:c.*667A>C
NM_001145849.1:c.*667A>C
NM_001145850.1:c.*667A>C
NM_001145851.1:c.*667A>C
NM_001145852.1:c.*667A>C
NM_006017.2:c.*667A>C

In Ensembl Biomart

NM_001145847
>ENSG00000007062|ENST00000508167|PROM1 Sequence of length 1168

NM_001145848

>ENSG00000007062|ENST00000505450|PROM1 Sequence of length 1165

NM_001145849

>ENSG00000007062|ENST00000540805|PROM1

TAGCTGATGTTGAAACTGCTTGAGCATCAGGATACTCAAAGTGGAAAGGATCACAGATTTTTGGTAGTTTCTGGGTCTACAAGGACTTTCCAAA
TCCAGGAGCAACGCCAGTGGCAACGTAGTGACTCAGGCGGGCACCAAGGCAACGGCACCATTGGTCTCTGGGTAGTGCTTTAAGAATGAACACA
ATCACGTTATAGTCCATGGTCCATCACTATTCAAGGATGACTCCCTCCCTTCCTGTCTATTTTTGTTTTTTACTTTTTTACACTGAGTTTCTAT
TTAGACACTACAACATATGGGGTGTTTGTTCCCATTGGATGCATTTCTATCAAAACTCTATCAAATGTGATGGCTAGATTCTAACATATTGCCA
TGTGTGGAGTGTGCTGAACACACACCAGTTTACAGGAAAGATGCATTTTGTGTACAGTAAACGGTGTATATACCTTTTGTTACCACAGAGTTTT
TTAAACAAATGAGTATTATAGGACTTTCTTCTAAATGAGCTAAATAAGTCACCATTGACTTCTTGGTGCTGTTGAAAATAATCCATTTTCACTA
AAAGTGTGTGAAACCTACAGCATATTCTTCACGCAGAGATTTTCATCTATTATACTTTATCAAAGATTGGCCATGTTCCACTTGGAAATGGCAT
GCAAAAGC[A/C]ATCATAGAGAAACCTGCGTAACTCCATCTGACAAATTCAAAAGAGAGAGAGATCTTGAGAGAGAAATGCTGTTCGTTCA
AAAGTGGAGTTGTTTTAACAGATGCCAATTACGGTGTACAGTTTAACAGAGTTTTCTGTTGCATTACGATAAACATTAATTGGAGTGCAGCTAA
CATGAGTATCATCAGACTAGTATCAAGTGTTCTAAAATGAAATATGAGAAGATCCTGTCACAATTCTTAGATCTGGTGTCCAGCATGGATGAAA
CCTTTGAGTTTGGTCCCTAAATTTGCATGAAAGCACAAGGTAAATATTCATTTGCTTCAGGAGTTTCATGTTGGATCTGTCATTATCAAAAGTG
ATCAGCAATGAAGAACTGGTCGGACAAAATTTAACGTTGATGTAATGAAATTCCAGATGTAGGCATTCCCCCCAGGTCTTTTCATGTGCAGATT
GCAGTTCTGATTCATTTGAATAAAAAGGAACTTGGAAAACATG

NM_001145850

>ENSG00000007062|ENST00000539194|PROM1 Same sequence of length 1167

NM_001145851

>ENSG00000007062|ENST00000540805|PROM1 Same sequence of length 1167

NM_001145852

>ENSG00000007062|ENST00000539194|PROM1 Same sequence of length 1167

NM_006017

>ENSG00000007062|ENST00000447510|PROM1 Same sequence of length 1167

Other available 3'UTR transcripts in Biomart:

>ENSG00000007062|ENST00000503884|PROM1 Sequence of length 425
>ENSG00000007062|ENST00000513946|PROM1 Sequence of length 301
>ENSG00000007062|ENST00000511153|PROM1 Sequence of length 1432
>ENSG00000007062|ENST00000510224|PROM1 Sequence of length 1159

Average Sequence length: 1044

Unavailable Sequences : ENST00000514693, ENST00000513108, ENST00000511270, ENST00000513448, ENST00000508940, ENST00000504842, ENST00000508322, ENST00000514967, ENST00000502943, ENST00000502501, ENST00000512304

PDCD1

L(3'UTR) nt	1177
mRNA RefSeq No	NM_005018
Ensembl Gene ID	ENSG00000276977
Ensembl Transcript ID	ENST00000618185
SNP ID	rs10204525

In publication

rs10204525 (+8669 A/G) SNP falls within the seed region of the target site (showed in a figure).

In dbSNP

AGTGCAGGCACCTAGGGCCCCCCAT[A/G]TGCCCACCCTGGGAGCTCTCCTTGG

NM_005018.2:c.*889G>A

In Ensembl Biomart

NM_005018

>ENSG00000188389|ENST00000334409|PDCD1
>ENSG00000276977|ENST00000618185|PDCD1
CCGGCTTCCTTGGCCACCAGTGTTCTGCAGACCCTCCACCATGAGCCCGGGTCAGCGCATTTCCTCAGGAGAAGCAGGCAGGGTGCAGGCCATT
GCAGGCCGTCCAGGGGCTGAGCTGCCTGGGGGCGACCGGGGCTCCAGCCTGCACCTGCACCAGGCACAGCCCCACCACAGGACTCATGTCTCAA
TGCCCACAGTGAGCCCAGGCAGCAGGTGTCACCGTCCCCTACAGGGAGGGCCAGATGCAGTCACTGCTTCAGGTCCTGCCAGCACAGAGCTGCC
TGCGTCCAGCTCCCTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCCTGCGGCCCGGGGCTGAAGGCGCCGTGGCCCTGCCTGACGCC
CCGGAGCCTCCTGCCTGAACTTGGGGGCTGGTTGGAGATGGCCTTGGAGCAGCCAAGGTGCCCCTGGCAGTGGCATCCCGAAACGCCCTGGACG
CAGGGCCCAAGACTGGGCACAGGAGTGGGAGGTACATGGGGCTGGGGACTCCCCAGGAGTTATCTGCTCCCTGCAGGCCTAGAGAAGTTTCAGG
GAAGGTCAGAAGAGCTCCTGGCTGTGGTGGGCAGGGCAGGAAACCCCTCCACCTTTACACATGCCCAGGCAGCACCTCAGGCCCTTTGTGGGGC
AGGGAAGCTGAGGCAGTAAGCGGGCAGGCAGAGCTGGAGGCCTTTCAGGCCCAGCCAGCACTCTGGCCTCCTGCCGCCGCATTCCACCCCAGCC
CCTCACACCACTCGGGAGAGGGACATCCTACGGTCCCAAGGTCAGGAGGGCAGGGCTGGGGTTGACTCAGGCCCCTCCCAGCTGTGGCCACCTG
GGTGTTGGGAGGGCAGAAGTGCAGGCACCTAGGGCCCCCCAT[A/G]TGCCCACCCTGGGAGCTCTCCTTGGAACCCATTCCTGAAATTATTTA
AAGGGGTTGGCCGGGCTCCCACCAGGGCCTGGGTGGGAAGGTACAGGCGTTCCCCCGGGGCCTAGTACCCCCGCCGTGGCCTATCCACTCCTCA
CATCCACACACTGCACCCCCACTCCTGGGGCAGGGCCACCAGCATCCAGGCGGCCAGCAGGCACCTGAGTGGCTGGGACAAGGGATCCCCCTTC
CCTGTGGTTCTATTATATTATAATTATAATTAAATATGAGAGCATGCTAAGGA

Other available 3'UTR transcripts in Biomart:

>ENSG00000188389|ENST00000418831|PDCD1 Sequence of length 791
>ENSG00000276977|ENST00000630560|PDCD1 Sequence of length 791

Average Sequence length: 984

Unavailable Sequences : ENST00000343705, ENST00000630230

.2 Site Accessibility compared with the Conservation Feature in MicroRNA Target prediction

This analysis was carried out in early 2015. The STarMir and PITA web servers were used to generate predictions and the TargetScan ranked lists were sourced from release was 6.2. In the following tables of ranked lists, the highlighted genes are experimentally verified targets. The interaction scores computed by STarMir and PITA are given in parentheses. The last table gives the ranking score of each method by summing up its scores for all miRNAs.

MicroRNAs from our dataset

hsa-miR-221-3p

	Targetscan 6.2	STarMir	PITA
SNX4(1135)	1	4(-188.291)	5(-7.59)
RGS6(4063)	2	1(-920.254)	1(-19.00)
CDKN1B (1344)	3	5(-145.344)	4(-8.71)
CXCL12(3163)	4	3(-585.671)	3(-12.67)
TCF12(2371)	5	2(-589.71)	2(-15.96)

hsa-let-7e-5p

	Targetscan 6.2	STarMir	PITA
C14orf28(1707)	1	4(-410.039)	3(-15.78)
FIGNL2(2560)	2	3 (-489.319)	2(-16.07)
HMGA2(2999)	3	1 (-1011.974)	1(-17.31)
LIN28B(4548)	4	2(-835.916)	5(-12.97)
TRIM71(468)	5	5(-111.959)	4(-13.54)

hsa-miR-7-5p

	Targetscan 6.2	STarMir	PITA
UBXN2B(3950)	1	1(-480.296)	1(-12.16)
SPATA2(2228)	2	4(-137.187)	5(-11.27)
C5orf22(2156)	3	2(-299.734)	3(-11.36)
ZNF828(1047)	4	6(-128.544)	2(-11.53)
POLE4(706)	5	8(-72.345)	8(-9.57)
CNO(797)	6	10(-47.391)	9(-9.03)
RAF1(913)	7	9(-55.174)	7(-10.43)
IDE(2760)	8	3(-295.805)	4(-11.34)
EGFR(1721)	9	5(-131.624)	6(-10.58)
RELA(782)	10	7(-73.938)	10(-6.21)

hsa-miR-155-5p

	Targetscan 6.2	STarMir	PITA
QKI(5435)	1	2(-443.291)	3(-8.26)
ZBTB38(3814)	2	3(-383.593)	2(-9.77)
JARID2(1791)	3	4(-166.433)	4(-7.69)
RAB11FIP2(4082)	4	1(-484.856)	1(-10.7)
C2orf80(460)	5	5(-34.778)	5(-5.49)

hsa-miR-122-5p

	Targetscan 6.2	STarMir	PITA
CLIC4(3387)	1	1(-669.789)	7(-10.80)
CTDNEP1(598)	2	10(-82.92)	6(-12.41)
MASP1(1602)	3	3(-384.86)	3(-14.57)
LMNB2(2745)	4	2(-445.015)	1(-16.33)
SPOCK2(3726)	5	4(-338.386)	4(-12.48)
PRKRA(683)	6	8(-93.784)	5(-12.44)
RFXAP(1850)	7	5(-246.67)	9(-8.58)
P4HA1(1000)	8	9(-85.602)	10(-8.20)
MAP3K12(690)	9	6(-174.154)	8(-10.79)
NICN1(2500)	10	7(-150.57)	2(-14.72)

hsa-miR-17-5p

	Targetscan 6.2	STarMir	PITA
FGD4(5518)	1	2(-792.261)	2(-12.62)
PKD2(2085)	2	3(-318.783)	5(-11.82)
MAP3K2(8910)	3	1(-1359.315)	4(-12.24)
ZNFX1(1365)	4	4(-315.556)	1(-21.77)
PDCD1LG2(1323)	5	5(-305.503)	3(-12.44)

hsa-miR-151a-5p

	Targetscan 6.2	STarMir	PITA
WNT1(973)	1	5(-256.451)	3(-11.61)
SEZ6L(3302)	2	3(-596.447)	2(-11.65)
NTRK2(6081)	3	1(-891.66)	4(-11.33)
N4BP1(4178)	4	4(-569.642)	1(-27.63)
PHF15(3911)	5	2(-606.699)	5(-10)

hsa-miR-27b-3p

	Targetscan 6.2	STarMir	PITA
C1orf144(3002)	1	2(-323.552)	1(-16.12)
ONECUT2(14575)	2	1(-883.886)	3(-11.57)
SEMA6A(3058)	3	5(-225.827)	2(-12.13)
MED12L(3873)	4	3(-306.534)	4(-10.53)
DCUN1D4(3253)	5	4(-280.19)	5(-8.62)

Let7e family

hsa-miR-98-5p

	Targetscan 6.2	STarMir	PITA
C14orf28(1707)	1	3(-235.682)	3(-13.80)
FIGNL2(2560)	2	4(-143.786)	4(-9.83)
HMGA2(2999)	3	2(-501.509)	1(-16.43)
LIN28B(4548)	4	1(-512.255)	2(-14.04)

TRIM71(468)	5	5(-50.402)	5(-9.81)

hsa-let-7a-5p

	Targetscan 6.2	**STarMir**	**PITA**
C14orf28(1707)	1	3(-358.25)	4(-13.02)
FIGNL2(2560)	2	4(-288.242)	3(-14.74)
HMGA2(2999)	3	1(-767.099)	1(-18.92)
LIN28B(4548)	4	2(-642.169)	2(-14.98)
TRIM71(468)	5	5(-100.548)	5(-10.6)

hsa-let-7b-5p

	Targetscan 6.2	**STarMir**	**PITA**
C14orf28(1707)	1	4(-383.811)	4(-13.47)
FIGNL2(2560)	2	3(-402.358)	2(-16.26)
HMGA2(2999)	3	1(-1042.27)	1(-19.98)
LIN28B(4548)	4	2(-1060.15)	3(-15.31)
TRIM71(468)	5	5(-100.829)	5(-11.27)

hsa-let-7c-5p

	Targetscan 6.2	**STarMir**	**PITA**
C14orf28(1707)	1	3(-324.371)	4(-13.12)
FIGNL2(2560)	2	4(-266.478)	2(-16.43)
HMGA2(2999)	3	2(-824.512)	1(-18.85)
LIN28B(4548)	4	1(-840.032)	3(-15.44)
TRIM71(468)	5	5(-85.116)	5(-11.33)

hsa-let-7d-5p

	Targetscan 6.2	**STarMir**	**PITA**
C14orf28(1707)	1	4(-389.889)	5(-11.7)
FIGNL2(2560)	2	3(-414.082)	2(-17)
HMGA2(2999)	3	2(-796.627)	1(-18.3)
LIN28B(4548)	4	1(-867.816)	3(-15.67)
TRIM71(468)	5	5(-114.416)	4(-12.1)

hsa-let-7g-5p

	Targetscan 6.2	**STarMir**	**PITA**
C14orf28(1707)	1	3(-325.43)	4(-12.26)
FIGNL2(2560)	2	4(-238.759)	3(-14.71)
HMGA2(2999)	3	2(-550.577)	1(-21.07)
LIN28B(4548)	4	1(-572.898)	2(-16.18)
TRIM71(468)	5	5(-91.508)	5(-11.09)

Mir 10a/b family

hsa-mir-10b

	Targetscan 6.2	STarMir	PITA
ARSJ(1990)	1	7(-66.58)	8(-5.66)
CADM2(7646)	2	1(-202.453)	2(-14.28)
SOBP(3106)	3	4(-121.375)	1(-15.70)
BDNF(2926)	4	5(-110.456)	5(-9.08)
FIGN(1944)	5	10(-30.673)	10(-4.26)
KLHL29(1949)	6	3(-137.514)	9(-5.43)
TFAP2C(1269)	7	6(-71.959)	3(-13.51)
CRLF3(1503)	8	9(-44.147)	6(-8.71)
KLHDC10(4974)	9	2(-173.597)	7(-8.01)
NCOR2(991)	10	8(-61.828)	4(-13.38)

Mir 24 family

hsa-mir-24

	Targetscan 6.2	STarMir	PITA
IFFO2(4281)	1	4(-623.7)	1(-16.43)
BCL2L11(4214)	2	3(-672.746)	3(-13.36)
DNAJB12(2980)	3	5(-547.801)	4(-13.06)
TAOK1(8881)	4	1(-1075.51)	5(-12.88)
NDST1(4879)	5	2(-781.85)	2(-15.19)

Mir 25 family

hsa-mir-92a

	Targetscan 6.2	STarMir	PITA
CD69(995)	1	4(-90.509)	2(-11.57)
FNIP1(2971)	2	3(-194.713)	3(-11.19)
SLC12A5(2545)	3	1(-346.633)	5(-8.18)
MAN2A1(730)	4	5(-50.252)	1(-12.40)
ACTC1(2316)	5	2(-308.603)	4(-10.26)

Mir 26 family

hsa-mir-26a

	Targetscan 6.2	STarMir	PITA
KLHDC5(4874)	1	1(-338.692)	1(-12.75)
TET2(3284)	2	2(-218.908)	5(-6.92)
STRADB(818)	3	5(-68.6)	2(-12.25)
CHORDC1(1976)	4	4(-96.853)	3(-8.74)
FAM98A(1114)	5	3(-113.707)	4(-7.42)

Mir 27 family

hsa-mir-27a

	Targetscan 6.2	STarMir	PITA
GXYLT1(5940)	1	1(-684.089)	1(-12.25)
PLK2(600)	2	10(-86.624)	4(-11.6)
AKIRIN1(2010)	3	2(-208.847)	3(-11.90)
GCC2(1768)	4	9(-99.906)	9(-7.91)
SBF2(1751)	5	6(-128.655)	2(-12.17)
RGPD4(1813)	6	8(-101.78)	8(-8.14)
RGPD8(1809)	7	7(-103.995)	7(-8.29)
PDS5B(2968)	8	3(-205.428)	10(-6.93)
CDS1(2654)	9	4(-139.597)	6(-9.22)
FBXW7(1623)	10	5(-134.736)	5(-9.88)

Mir 29abc family

hsa-mir-29a

	Targetscan 6.2	STarMir	PITA
ATAD2B(3382)	1	2(-176.597)	5(-6.77)
COL3A1(972)	2	5(-42.857)	3(-9.71)
ELN(1206)	3	3(-60.436)	2(-9.95)
HBP1(1098)	4	4(-50.966)	1(-10.03)
COL4A4(4614)	5	1(-193.804)	4(-8.66)

hsa-mir-29b

	Targetscan 6.2	STarMir	PITA
ATAD2B(3382)	1	3(-95.391)	5(-6.74)
COL3A1(972)	2	5(-49.498)	2(-10.05)
ELN(1206)	3	4(-64.484)	4(-8.93)
HBP1(1098)	4	2(-106.69)	1(-10.49)
COL4A4(4614)	5	1(-188.83)	3(-9.72)

hsa-mir-29c

	Targetscan 6.2	STarMir	PITA
ATAD2B(3382)	1	2(-101.918)	5(-6.89)
COL3A1(972)	2	4(-46.083)	1(-11.70)
ELN(1206)	3	5(-40.283)	2(-11.13)
HBP1(1098)	4	3(-53.723)	3(-10.03)
COL4A4(4614)	5	1(-210.915)	4(-8.99)

Mir 93/105 family

hsa-mir-302b

	Targetscan 6.2	STarMir	PITA

CROT(1152)	1	5(-74.518)	5(-5.97)
FGD4(5517)	2	1(-503.496)	2(-9.02)
LATS2(1833)	3	4(-117.18)	3(-7.21)
TGFBR2(2542)	4	3(-195.661)	4(-6.1)
ZKSCAN1(3583)	5	2(-269.55)	1(-10.54)

hsa-mir-372

	Targetscan 6.2	**STarMir**	**PITA**
CROT(1152)	1	5(-188.766)	5(-6.45)
FGD4(5517)	2	1(-522.948)	2(-13.33)
LATS2(1833)	3	4(-209.713)	4(-8.24)
TGFBR2(2542)	4	2(-365.351)	3(-11.24)
ZKSCAN1(3583)	5	3(-337.7)	1(-14.8)

hsa-mir-373

	Targetscan 6.2	**STarMir**	**PITA**
CROT(1152)	1	5(-168.31)	5(-8.94)
FGD4(5517)	2	1(-1215.109)	2(-15.21)
LATS2(1833)	3	4(-269.555)	4(-9.37)
TGFBR2(2542)	4	3(-766.765)	3(-11.48)
ZKSCAN1(3583)	5	2(-1011.351)	1(-15.71)

Mir 96 family

hsa-mir-96

	Targetscan 6.2	**STarMir**	**PITA**
KIAA2022(6549)	1	2(-446.166)	5(-9.56)
PRTG(8471)	2	1(-609.319)	2(-11.53)
ADCY6(2399)	3	4(-164.654)	4(-9.59)
FRS2(4827)	4	3(-322.267)	3(-10.76)
NLGN2(2061)	5	5(-97.449)	1(-13.56)

Mir 125b family

hsa-mir-125b

	Targetscan 6.2	**STarMir**	**PITA**
FAM169B(2469)	1	1(-160.629)	3(-11.30)
STARD13(2458)	2	3(-111.257)	4(-7.14)
ZNF792(1602)	3	2(-120.197)	5(-7.10)
CCR2(808)	4	4(-71.349)	2(-11.93)
ACHE(964)	5	5(-15.322)	1(-12.67)

hsa-mir-125a-5p

	Targetscan 6.2	STarMir	PITA
RORA(9171)	1	3(-525.238)	2(-15.09)
ARID3B(2357)	2	5(-134.965)	5(-12.03)
PRTG(8471)	3	1(-635.548)	4(-12.6)
ZNF704(12915)	4	2(-616.997)	1(-26.1)
ANKRD33B(7703)	5	4(-468.16)	2(-15.09)

Mir 126 family

hsa-mir-126

	Targetscan 6.2	STarMir	PITA
PTPN9(1648)	1	2(-73.181)	2(-9.38)
PLXNB2(725)	2	4(-23.993)	5(-0.96)
RGS3(785)	3	5(-14.33)	4(-1.03)
KANK2(2291)	4	1(-104.024)	1(-13.75)
EFHD2(1619)	5	3(-50.695)	3(-8.14)

Mir 135ab family

hsa-mir-135a

	Targetscan 6.2	STarMir	PITA
ANGPT2(3447)	1	4(-179.755)	3(-10.89)
GK5(8095)	2	1(-437.486)	2(-11.14)
NR3C2(2580)	3	3(-193.814)	1(-11.83)
GULP1(1914)	4	5(-98.218)	4(-9.86)
LOC221710(4134)	5	2(-236.434)	5(-8.37)

Mir 137 family

hsa-mir-137

	Targetscan 6.2	STarMir	PITA
PDLIM3(1633)	1	5(-56.714)	4(-5.86)
APPL2(1017)	2	2(-77.14)	5(-5.26)
MITF(3089)	3	3(-73.581)	3(-6.74)
CEP128(1001)	4	4(-71.649)	1(-9.06)
RAVER2(2246)	5	1(-77.128)	2(-7.77)

Mir 139-5p family

hsa-mir-139-5p

	Targetscan 6.2	STarMir	PITA
HNRNPU(4150)	1	4(-527.276)	1(-13.05)
MNT(2841)	2	5(-425.775)	4(-11.46)
PPARGC1B(7464)	3	2(-806.886)	5(-9.97)

IGF1R(7088)	4	1(-888.289)	3(-11.54)
KPNA4(7109)	5	3(-614.984)	2(-12.69)

Mir 140-5p family

hsa-mir-140-5p

	Targetscan 6.2	**STarMir**	**PITA**
ZNF800(1514)	1	4(-136.034)	3(-9.83)
YOD1(5171)	2	1(-434.662)	1(-11.33)
FGF9(3066)	3	2(-235.338)	5(-8.17)
SEPT2(2089)	4	3(-176.067)	2(-10.19)
MMD(1721)	5	5(-129.946)	4(-8.48)

Mir 141family

hsa-mir-141

	Targetscan 6.2	**STarMir**	**PITA**
ZFR(1410)	1	5(-81.363)	5(-6.43)
RANBP6(1271)	2	4(-84.99)	1(-15.51)
ZEB2(5076)	3	2(-375.203)	3(-9.4)
ABL2(8391)	4	1(-469.476)	2(-9.8)
ARPC5(1339)	5	3(-90.005)	4(-9.04)

hsa-mir-200a

	Targetscan 6.2	**STarMir**	**PITA**
ZFR(1410)	1	5(-79.015)	5(-5.75)
RANBP6(1271)	2	3(-109.951)	1(-12.31)
ZEB2(5076)	3	2(-254.345)	3(-8.19)
ABL2(8391)	4	1(-405.716)	2(-8.33)
ARPC5(1339)	5	4(-84.156)	4(-7.71)

Mir 143family

hsa-mir-143

	Targetscan 6.2	**STarMir**	**PITA**
GYGYF2(3719)	1	3(-295.717)	4(-10.95)
SH3PXD2A(7789)	2	1(-833.424)	5(-10.27)
KRAS(4549)	3	2(-456.871)	1(-14.3)
TPM3(1275)	4	5(-139.301)	2(-12.83)
ETV6(4345)	5	4(-273.459)	3(-11.3)

Results after computing the ranking score of each method by summing up its scores for all miRNAs

	Targetscan 6.2 Ranking score	STarMir Ranking score	PITA Ranking score
hsa-miR-221-3p	1	3	2
hsa-let-7e-5p	2	1	1
hsa-miR-7-5p	3	1	2
hsa-miR-155-5p	2	1	1
hsa-miR-122-5p	2	3	1
hsa-miR-17-5p	2	3	1
hsa-miR-151a-5p	2	2	1
hsa-miR-27b-3p	2	3	1
hsa-miR-98-5p	3	2	1
hsa-let-7a-5p	3	1	1
hsa-let-7b-5p	3	1	1
hsa-let-7c-5p	3	1	1
hsa-let-7d-5p	3	2	1
hsa-let-7g-5p	3	2	1
hsa-mir-10b	3	2	1
hsa-mir-24	1	2	2
hsa-mir-92a	1	2	2
hsa-mir-26a	2	3	1
hsa-mir-27a	3	1	2
hsa-mir-29a	2	2	1
hsa-mir-29b	1	3	1
hsa-mir-29c	2	3	1
hsa-mir-302b	2	1	2
hsa-mir-372	1	2	2
hsa-mir-373	1	2	2
hsa-mir-96	1	2	2
hsa-mir-125b	1	2	3
hsa-mir-125a-5p	1	2	2
hsa-mir-126	1	3	2
hsa-mir-135a	2	2	1
hsa-mir-137	1	1	1
hsa-mir-139-5p	3	1	2
hsa-mir-140-5p	2	1	3
hsa-mir-141	2	1	2
hsa-mir-200a	2	1	2
hsa-mir-143	3	2	1
Total	72	67	54

.3 Accessibility of microRNA binding sites in metastable RNA secondary structures in the presence of SNPs - PITA and STarMir predictions

LIG3; rs4796030; mir-221

A-Allele

STarMir predicts 13 seedless binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
9	77-93	No	-15	-3.297	-5.236	-9.764	0.669
10	77-96	No	-15.500	-3.383	-6.527	-8.973	0.628
11	77-98	No	-20.300	-3.388	-11.769	-8.531	0.577
12	81-107	No	-21.400	-0.513	-19.753	-1.647	0.460

```
Site 9

    5'->3'        G    AGA     CU   U
    Target    77   GAAA   CCAGU  GGG            93
                   ||||   |||||  |||
    miRNA     23   CUUU   GGUCG  UCU             1
    3'->5'             G            GUUACAUCGA


Site 10

    5'->3'        G    AGA     CU        N
    Target    77   GAAA   CCAGU  GG  GUGU      96
                   ||||   |||||  ||  ||||
    miRNA     23   CUUU   GGUCG  CU  UACA      1
    3'->5'             G       U   GU    UCGA


Site 11

    5'->3'        G    AGA     CUG        G
    Target    77   GAAA   CCAGU    GGUGUGG    98
                   ||||   |||||    |||||||
    miRNA     23   CUUU   GGUCG    UUACAUC    1
    3'->5'             G       UCUG       GA

The bindings in green here highlight positions predicted by
PITA (94-99)

Site 12

    5'->3'        A     AGUCU  GU UG GA   C   A
    Target    81   AGACC     GG  G  G  AUG AGC   107
                   |||||     ||  |  |  ||| |||
    miRNA     23   UUUGG     UC  C  U  UAC UCG   1
    3'->5'        C     G      GU UG      A   A
```

PITA predicts 2 sites with an overall score -5.98.
There is no binding site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-221	94-99	6:1:1	-15.6	-9.61	-5.98

C-Allele

STarMir predicts 14 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
9	80-93	No	-19.300	-0.131	-8.888	-10.412	0.492
10	80-96	No	-19.800	-0.176	-9.749	-10.051	0.490
11	80-88	No	-18	-0.131	-8.513	-9.487	0.549
12	80-98	No	-24.600	-0.197	-15.961	-8.639	0.441
13	80-107	No	-25.500	-0.160	-23.586	-1.914	0.406

```
Site 9

    5'->3'        A         CU   U
    Target    80   AAGCCCAGU  GGG            93
                   |||||||||  |||
    miRNA     23   UUUGGGUCG  UCU             1
    3'->5'        C              GUUACAUCGA


Site 10

    5'->3'        A         CU        N
    Target    80   AAGCCCAGU  GG  GUGU      96
                   |||||||||  ||  ||||
    miRNA     23   UUUGGGUCG  CU  UACA      1
    3'->5'        C         U   GU    UCGA


Site 11

    5'->3'        A         C
    Target    80   AAGCCCAGU               88
                   |||||||||
    miRNA     23   UUUGGGUCG               1
    3'->5'        C         UCUGUUACAUCGA


Site 12

    5'->3'        A         CUG        G
    Target    80   AAGCCCAGU    GGUGUGG    98
                   |||||||||    |||||||
    miRNA     23   UUUGGGUCG    UUACAUC    1
    3'->5'        C         UCUG       GA


The bindings in green highlight positions predicted by PITA
(94-99)
```

PITA predicts 2 sites with an overall score -7.45.
There is no binding site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	hsa-miR-221	94-99	6:1:1	-19.8	-12.34	-7.45

CBR1; rs9024; mir-574-5p

G-Allele

STarMir predicts 15 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
10	121-162	No	-27.800	-5.364	-8.085	-19.715	0.527
11	121-137	No	-21.300	-3.355	-3.283	-18.017	0.528
12	121-140	No	-21.600	-3.538	-4.309	-17.291	0.565

```
Site 10

    5'->3'         A        UAAUGUAC   UAAUUGAGCAACCU         G
    Target   121    GCACUCAC         UAC               ACGCACUCA   162
                    ||||||||         |||               |||||||||
    miRNA     23    UGUGAGUG         GUG               UGUGUGAGU   1
    3'->5'         UG        U

The bindings in green highlight positions predicted by PITA
154-161 with a perfect 8 mer with one G-U wobble


Site 11

    5'->3'         A        UA         A
    Target   121    GCACUCAC    AUGUACU    137
                    ||||||||    |||||||
    miRNA     23    UGUGAGUG    UGUGUGA    1
    3'->5'         UG        UGUG       GU


Site 12

    5'->3'         A        UAAUG        A
    Target   121    GCACUCAC      UAC UACU    140
                    ||||||||      ||| ||||
    miRNA     23    UGUGAGUG      GUG GUGA    1
    3'->5'         UG        UGU     U    GU
```

PITA predicts 3 sites with an overall score -17.94.
There is no binding site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	hsa-miR-574-5p	154-161	8:0:1	-24.49	-6.54	-17.94

A-Allele

STarMir predicts 16 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
10	121-137	No	-25.300	-4.513	-2.874	-22.426	0.584
11	121-140	No	-25.500	-4.599	-3.514	-21.986	0.631
12	121-162	No	-27.800	-5.224	-5.958	-21.842	0.563
13	121-135	No	-20.100	-3.708	-3.88	-16.220	0.574

```
Site 10

    5'->3'         A        UA         A
    Target   121    GCACUCAC    AUAUACU    137
                    ||||||||    |||||||
    miRNA     23    UGUGAGUG    UGUGUGA    1
    3'->5'         UG        UGUG       GU


Site 11

    5'->3'         A        UA          A
    Target   121    GCACUCAC  AUAUAC UACU    140
                    ||||||||  |||||| ||||
    miRNA     23    UGUGAGUG  UGUGUG GUGA    1
    3'->5'         UG                U    GU


Site 12

    5'->3'         A        UAAUAUAC   UAAUUGAGCAACCU         G
    Target   121    GCACUCAC         UAC               ACGCACUCA   162
                    ||||||||         |||               |||||||||
    miRNA     23    UGUGAGUG         GUG               UGUGUGAGU   1
    3'->5'         UG        U

The bindings in green highlight positions predicted by PITA
154-161 with a perfect 8 mer with one G-U wobble


Site 13

    5'->3'         A        UA       N
    Target   121    GCACUCAC    AUAUA      135
                    ||||||||    |||||
    miRNA     23    UGUGAGUG    UGUGU      1
    3'->5'         UG        UGUG     GAGU
```

PITA predicts 4 sites with an overall score -19.21.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-574-5p	133-138	6:1:1	-19.8	-2.56	-17.23

The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-574-5p	154 - 161	8:0:1	-24.49	-5.41	-19.07

FindTar predicts 3 binding sites for each sequence. Only the A-Allele has one site covering the SNP position:

hsa-miR-574-5p	A-Allele	133-162	3' UGUG-UGA-----GUGU-GUGUGUGUGAGU 5' |:||*|||******:||***:||:|||||* 5' ATACTACTAATTGAGCAACCTACGCACTCA 3'	25.00	-22.40	excellent

HTR3E; rs56109847; miR-510-5p

G-Allele

STarMir predicts 16 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
6	50-80	73-79	-28.700	-5.505	-12.645	-16.055	0.544
7	50-76	No	-20.300	-6.324	-11.509	-8.791	0.580

```
Site 6

    5'->3'          G  CUG    GGUCUCCCCC            C
    Target    50     GA   GCCA           CUUUCCUGAGUA   80
                     ||   ||||           ||||||||||||
    miRNA     22     CU   CGGU           GAGAGGACUCAU   1
    3'->5'         CA  AA

The bindings in green highlight positions predicted by PITA
72-79 with a perfect 8 mer.

Site 7

    5'->3'          G  CUG    GGUCUCCCCC        N
    Target    50     GA   GCCA           CUUUCCUG      76
                     ||   ||||           ||||||||
    miRNA     22     CU   CGGU           GAGAGGAC      1
    3'->5'         CA  AA                      UCAU
```

PITA predicts 6 sites with an overall score -16.41.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	miR-510-5p	72-79	8:0:0	-22.5	-6.08	-16.41

FindTar predictions

Position	Structure	Loop Score	ΔG
60-80	3' CACUAACGGUGAGAGGACUCAU 5' *******||*||:||||||||| 5' GTCTCCCCC-CTTTCCTGAGTA 3'	10.00	-24.90

A-Allele

STarMir predicts 16 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
6	50-80	No	-22.500	-2.464	-8.261	-14.239	0.566

```
Site 6

    5'->3'          G  CUG    GGUCUCCCCC        A    C
    Target    50     GA   GCCA           CUUUCCU AGUA   80
                     ||   ||||           ||||||| ||||
    miRNA     22     CU   CGGU           GAGAGGA UCAU   1
    3'->5'         CA  AA                      C

The bindings in green highlight positions predicted by PITA
72-79 with a 8 mer with one mismatch
```

PITA predicts 6 sites with an overall score -12.28.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	miR-510-5p	72-79	8:1:0	-16.3	-4.01	-12.28

HLA_G; rs1063320; mir-148a-3p

C-Allele

STarMir predicts 6 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
4	221-239	No	-23.900	-4.554	-9.282	-14.618	0.492
5	221-237	No	-18.900	-4.182	-9.628	-9.272	0.475
6	230-260	No	-9.800	-3.556	-9.067	-0.733	0.528

```
STarMir Site 4

    5'->3'        U              C      G
    Target   221   CAAA  UUUGUGGU CACUGA   239
                   ||||  |||||||| ||||||
    miRNA     22   GUUU  AGACAUCA GUGACU   1
    3'->5'        U    CA        C

The bindings in green highlight positions predicted by PITA
(231-238 with one mismatch)

STarMir Site 5

    5'->3'        U              C    N
    Target   221   CAAA  UUUGUGGU CACU    237
                   ||||  |||||||| ||||
    miRNA     22   GUUU  AGACAUCA GUGA    1
    3'->5'        U    CA        C    CU


STarMir Site 6

    5'->3'           U   CCA      CUAUAACUUACUUC      A
    Target   230      GGU   CUG AG               UGUAUU    260
                      |||   ||| ||               ||||||
    miRNA     22      UCA   GAC UC               ACGUGA    1
    3'->5'        UGUU   A     A                      CU
```

PITA predicts 7 sites with an overall score -1.16.
There is only one site covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	hsa-miR-148a	231-238	8:1:0	-16.8	-15.78	-1.01

FindTar predictions

220-239	3' UGUUUCAAGACAUCACGUGACU 5' *||||**|:|||:||*|||||| 5' TCAAA--TTTGTGGTCCACTGA 3'	-21.60

G-Allele

STarMir predicts 7 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
4	221-239	232-238	-30.500	-3.238	-11.006	-19.494	0.395
5	221-235	No	-21.100	-1.928	-9.059	-12.041	0.338
6	221-233	No	-15.600	-1.501	-11.349	-4.251	0.339

```
STarMir Site 4 (seed site)

    5'->3'        U                    G
    Target   221   CAAA  UUUGUGGUGCACUGA   239
                   ||||  ||||||||||||||||
    miRNA     22   GUUU  AGACAUCACGUGACU   1
    3'->5'        U    CA

The bindings in green highlight positions predicted by PITA.(
231-238 with perfect seed match)

STarMir Site 5

    5'->3'        U                  N
    Target   221   CAAA  UUUGUGGUGCA      235
                   ||||  ||||||||||||
    miRNA     22   GUUU  AGACAUCACGU      1
    3'->5'        U    CA           GACU


STarMir Site 6

    5'->3'        U                N
    Target   221   CAAA  UUUGUGGUG         233
                   ||||  |||||||||
    miRNA     22   GUUU  AGACAUCAC         1
    3'->5'        U    CA          GUGACU
```

PITA predicts 7 sites with an overall score -6.49.
There is only one site covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	hsa-miR-148a	231-238	8:0:0	-23.4	-16.90	-6.49

220-239	3' UGUUUCAAGACAUCACGUGACU 5' *||||**|:|||:|||||||| 5' TCAAA--TTTGTGGTGCACTGA 3'	-28.20

PARP1; rs8679; mir-145-5p

T-Allele

STarMir predicts 25 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
22	592-611	No	-18.300	-0.927	-10.683	-7.617	0.476
23	592-614	No	-19.700	-0.927	-12.191	-7.509	0.517

```
Site 22

    5'->3'            U     UCUCCA    U    A
    Target   592       UUCCU       GGAA ACUG      611
                       |||||       |||| ||||
    miRNA     23       AAGGA       CCUU UGAC      1
    3'->5'         UCCCU     C          U    CUG


Site 23

    5'->3'            U     UCUCCA    U    A  A
    Target   592       UUCCU       GGAA ACUG AC   614
                       |||||       |||| |||| ||
    miRNA     23       AAGGA       CCUU UGAC UG   1
    3'->5'         UCCCU     C          U    C

The bindings in green highlight positions predicted by PITA
608-614 with a 6 seed match with one mismatch
```

PITA predicts 7 sites with an overall score -2.53.
There is no binding site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
T-Allele	hsa-miR-145-5p	608-614	6:1:0	-12.8	-13.30	0.50
T-Allele	hsa-miR-145-5p	615-621	6:1:1	-8.64	-12.82	4.18

C-Allele

STarMir predicts 25 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
22	592-611	No	-17.800	-0.652	-17.666	-0.134	0.330
23	592-614	No	-19.200	-0.652	-20.904	1.704	0.347

```
Site 22

    5'->3'            U     UCUCCA    C    A
    Target   592       UUCCU       GGAA ACUG      611
                       |||||       |||| ||||
    miRNA     23       AAGGA       CCUU UGAC      1
    3'->5'         UCCCU     C          U    CUG


Site 23

    5'->3'            U     UCUCCA    C    A  A
    Target   592       UUCCU       GGAA ACUG AC   614
                       |||||       |||| |||| ||
    miRNA     23       AAGGA       CCUU UGAC UG   1
    3'->5'         UCCCU     C          U    C

The bindings in green highlight positions predicted by PITA
608-614 with a 6 seed match with one mismatch
```

PITA predicts 7 sites with an overall score -2.51.
There is no binding site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	hsa-miR-145-5p	608-614	6:1:0	-12.6	-13.33	0.73
C-Allele	hsa-miR-145-5p	615-621	6:1:1	-8.64	-12.80	4.16

WFS1; rs1046322; hsa-miR-668-3p

G-Allele

STarMir predicts 75 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
25	234-258	251-257	-26.600	-4.421	-10.641	-15.959	0.498

```
Site 25

    5'->3'              C       UGACCUUUCU        U
    Target   234         CUGAGCC           GAGUGACA   258
                         |||||||           ||||||||
    miRNA     23         GGCUCGG           CUCACUGU   1
    3'->5'        CAUCACCC

The bindings in green highlight positions predicted by PITA
251-257 with a perfect 7 mer.
```

PITA predicts 15 sites with an overall score -16.47.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	hsa-miR-668-3p	251	7:0:0	-24.05	-7.57	-16.47

FindTar predictions

Position	Structure	Loop Score	ΔG
236-258	3' CAUCACCCGGCUCGG----------CUCACUGU 5' ********|:||||**********|||||||* 5' TCTCCACCCTGAGCCTGACCTTTCTGAGTGACA 3'	20.00	-25.60

A-Allele

STarMir predicts 75 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
25	234-258	No	-20.600	-1.723	-14.196	-6.404	0.467

```
Site 25

    5'->3'              C       UGACCUUUCU  A     U
    Target   234         CUGAGCC           GA UGACA   258
                         |||||||           || |||||
    miRNA     23         GGCUCGG           CU ACUGU   1
    3'->5'        CAUCACCC                    C

The bindings in green highlight positions predicted by PITA
251-257 with a 7 mer with one mismatch
```

PITA predicts 15 sites with an overall score -11.13.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-668-3p	251	7:1:0	-18.05	-6.92	-11.12

IL23R; rs10889677; let-7e

C-Allele

STarMir predicts 39 seedless binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
13	291-310	No	-24.700	0	-25.819	1.119	0.201

```
Site 13

    5'->3'          U   GC  U            U
    Target   291     UUA  CA UCUUCUGCCUCA   310
                     |||  || ||||||||||||
    miRNA     22     GAU  GU GGAGGAUGGAGU   1
    3'->5'          UU   AU  U
```

The bindings in green highlight positions predicted by PITA 303 -309 (7 mer with one G-U wobble)

PITA predicts 14 sites with an overall score -14.23.
There is one binding site covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	hsa-let-7e	303 -309	7:0:1	-19.4	-5.27	-14.12

FindTar predictions

Position	Structure	Loop Score	ΔG
287-310	3' UUGAUAUGUUGGAGGAUGGAGU 5' **:||**||**||:||:||||* 5' TTTTAGCCATTCTTCTGCCTCA 3'	20.00	-23.50

A-Allele

STarMir predicts 38 binding sites.
There is no binding site covering the SNP position.

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
13	291-308	No	-20.700	-0.045	-21.396	0.696	0.205

```
Site 13

    5'->3'          U   GC  U            A
    Target   291     UUA  CA UCUUCUGCCU    308
                     |||  || ||||||||||
    miRNA     22     GAU  GU GGAGGAUGGA    1
    3'->5'          UU   AU  U            GU
```

The bindings in green highlight positions predicted by PITA 303 -309 (7 mer with one G-U wobble and one mismatch)

PITA predicts 14 sites with an overall score -12.11.
There is one binding site covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-let-7e	303-309	7:1:1	-15.9	-5.87	-10.02

Position	Structure	Loop Score	ΔG
287-310	3' UUGAUAUGUUGGAGGAUGGAGU 5' **:||**||**||:||:|||** 5' TTTTAGCCATTCTTCTGCCTAA 3'	20.00	-20.00

RYR3 ; rs1044129 ; miR-367

A-Allele

STarMir predicts 12 seedless binding sites and one seed site.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
11	830-847	No	-16.300	-0.335	-7.633	-8.667	0.582
12	835-857	852-857	-14.800	-1.222	-11.419	-3.381	0.532

```
STarMir Site 11

    5'->3'        C   AAUACAA        C
    Target   830   UCA        UGAAGUGC      847
                   |||        ||||||||
    miRNA     22   AGU        AUUUCACG      1
    3'->5'           GGUAACG         UUAA


Site 12 (seed site)
The bindings in green highlight positions predicted by PITA
but with A-U binding at the beginning (850-857)

    5'->3'        A   AAUGAAG    CCAC       A
    Target   835   UAC        UGC     UGCAAU    857
                   |||        |||     ||||||
    miRNA     22   GUG        ACG     ACGUUA    1
    3'->5'        A   GUA       AUUUC       A
```

PITA predicts 29 binding sites with an overall score -6.97.
There is only one binding site covering the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-367	835-840	6:1:0	-1.77	-6.76	4.99

The following sites may cause the opening of the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-367	850-857	8:1:0	-8.9	-12.16	3.26
A-Allele	hsa-miR-367	840-845	6:1:1	-1.95	-9.23	7.28

G-Allele

STarMir predicts 12 seedless binding sites and one seed site.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
11	830-847	No	-16.300	-0.840	-11.825	-4.475	0.463
12	830-857	852-857	-15.300	-0.194	-19.650	4.350	0.397

```
Site 11

    5'->3'        C   AAUA  G           C
    Target   830   UCA    CA    UGAAGUGC      847
                   |||    ||    ||||||||
    miRNA     22   AGU    GU    AUUUCACG      1
    3'->5'           G     AACG        UUAA


Site 12 (seed site)
The bindings in green highlight positions predicted by PITA but
with A-U binding at the beginning (850-857)

    5'->3'        C   AAUA  GUGAAG    CCAC       A
    Target   830   UCA    CA      UGC     UGCAAU    857
                   |||    ||      |||     ||||||
    miRNA     22   AGU    GU      ACG     ACGUUA    1
    3'->5'           G     A         AUUUC       A
```

PITA predicts 29 binding sites with an overall score -6.97.
There are two binding sites covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	hsa-miR-367	838-845	8:1:1	-7.7	-11.38	3.68
G-Allele	hsa-miR-367	835-840	6:1:1	-1.77	-8.92	7.15

The following site may cause the opening of the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	hsa-miR-367	850-857	8:1:0	-8.73	-14.31	5.58

AGTR1; rs5186 ; miR-155-5p

A-Allele

STarMir predicts 21 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
2	57-90	83-89	-20.900	-0.577	-28.561	7.661	0.283
3	57-86	No	-15.700	-0.085	-27.289	11.589	0.248

```
Site 2

    5'->3'        U      CAGCACU    UACCAAAUG        C
    Target    57   CCUCUG        UCAC          AGCAUUAG   90
                   ||||||        ||||          ||||||||
    miRNA     23   GGGGAU        AGUG          UCGUAAUU   1
    3'->5'        U                  CUAA

The bindings in green highlight positions predicted by PITA
83-89 with a perfect 7 mer.

Site 3

    5'->3'        U      CAGCACU    UACCAA  G    N
    Target    57   CCUCUG        UCAC       AU AGCA      86
                   ||||||        ||||       || ||||
    miRNA     23   GGGGAU        AGUG       UA UCGU      1
    3'->5'        U                  C        A    AAUU
```

PITA predicts 15 sites with an overall score -5.40.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	miR-155-5p	83-89	7:0:0	-14.27	-9.97	-4.29

The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	miR-155-5p	100	7:1:1	-2.12	-10.04	7.92

FindTar predictions

Position	Structure	Loop Score	ΔG
67-90	3' UGGGGAUAGUGCUAAUCGUAAUU 5' ****|||*||*****|||||||* 5' TTCACTACCA-AATGAGCATTAG 3'	20.00	-15.50

C-Allele

STarMir predicts 20 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
2	57-90	No	-16.600	-2.268	-27.879	11.279	0.284
3	84-123	118-123	-14.900	-0.429	-22.035	7.135	0.475

```
Site 2

    5'->3'        U      CAGCACU    UACCAAAUG   C    C
    Target    57   CCUCUG        UCAC          AGC UUAG   90
                   ||||||        ||||          ||| ||||
    miRNA     23   GGGGAU        AGUG          UCG AAUU   1
    3'->5'        U                  CUAA        U

The bindings in green highlight positions predicted by PITA
83-89 with a 7 mer with one mismatch


Site 3

    5'->3'        A    UAG   CUUU   GA    AAGGAGAAAAU      U
    Target    84   GCCU   CUA    UCA  AUUG           GCAUUA
123
                   ||||   |||    |||  ||||           ||||||
    miRNA     23   UGGG   GAU    AGU  UAAU           CGUAAU   1
    3'->5'                      GC                       U
```

PITA predicts 15 sites with an overall score -5.19.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	miR-155-5p	83-89	7:1:0	-9.97	-6.84	-3.12

The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	miR-155-5p	100	7:1:1	-7.7	-6.78	-0.91

FGF20; rs12720208; miR-433-3p

C-Allele

STarMir predicts 45 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
9	166-187	180-186	-14.500	-2.356	-9.274	-5.226	0.426

```
Site 9

    5'->3'          U    CU  AAAUAG        C
    Target   166     UUGA  AG      AUCAUGAU   187
                     ||||  ||      ||||||||
    miRNA     22     GGCU  UC      UAGUACUA   1
    3'->5'         UGU    CC  GGG


The bindings in green highlight positions predicted by PITA
180-186 with a perfect 7 mer.
```

PITA predicts 22 sites with an overall score -11.08.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	miR-433-3p	180-186	7:0:0	-10.55	-7.11	-3.43

FindTar predictions

Position	Structure	Loop Score	ΔG
163-187	3' UGUGGCUCCUC--GG-GUAGUACUA 5' ***::|||**||***:**||||||| 5' ATTTTGACTAGAAATAGATCATGAT 3'	20.00	-12.80

T-Allele

STarMir predicts 46 binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
9	166-187	No	-12.300	-4.655	-10.123	-2.177	0.547
10	166-185	No	-8.600	-3.477	-10.088	1.488	0.526

```
Site 9

    5'->3'          U    CU  AAAUAG        C
    Target   166     UUGA  AG      AUUAUGAU   187
                     ||||  ||      |||||||
    miRNA     22     GGCU  UC      UAGUACUA   1
    3'->5'         UGU    CC  GGG

The bindings in green highlight positions predicted by PITA
180-186 with a 7 mer with one G-U wobble
```

PITA predicts 22 sites with an overall score -11.08.
There is one binding site covering the SNP position.

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
T-Allele	miR-433-3p	180-186	7:0:1	-8.35	-6.70	-1.64

Position	Structure	Loop Score	ΔG
163-187	3' UGUGGCUCCUC--GG-GUAGUACUA 5' ***::||**||***:**||:||||| 5' ATTTTGACTAGAAATAGATTATGAT 3'	20.00	-10.60

HOXB5 ; rs9299 ; miR-7-5p

G-Allele

STarMir predicts 26 seedless binding sites and one seed site.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
1	126-145	No	-17.200	-0.428	-6.624	-10.576	0.363
2	126-154	148-154	-21.900	-0.493	-13.609	-8.291	0.389

```
STarMir Site 1

    5'->3'        U      UUUCGU        A
    Target   126   ACGAUA       UUGGUCUU     145
                   ||||||       ||||||||
    miRNA     23   UGUUGU       GAUCAGAA     1
    3'->5'               UUUAGU         GGU


STarMir Site 2 seed site

    5'->3'        U      UUUCGUUU    UUA        U
    Target   126   ACGAUA         GGUC   GGUCUUCC   154
                   ||||||         ||||   ||||||||
    miRNA     23   UGUUGU         UUAG   UCAGAAGG   1
    3'->5'               U            UGA        U

The bindings in green highlight positions predicted by PITA
(147-154)
```

PITA predicts 41 binding sites with an overall score -8.39.
There is no binding site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	hsa-miR-7	148-154	7:0:0	-16.4	-9.5	-6.89

FindTar predictions

Position	Structure	Loop Score	ΔG
126-155	3' UGUUGUUUUAGUGAUCAGAAGGU 5' *******::||***:||||||||* 5' TTCGTTTGGTCTTAGGTCTTCCT 3'	25.00	-20.80

A-Allele

STarMir predicts 26 seedless binding sites and one seed site.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
1	126-154	148-154	-22.400	-1.391	-11.198	-11.202	0.466
2	126-152	No	-15.100	-1.183	-9.139	-5.961	0.463

```
STarMir Site 1 seed site

    5'->3'        U      UUUCGUUU    UUA        U
    Target   126   ACGAUA         GAUC   GGUCUUCC   154
                   ||||||         ||||   ||||||||
    miRNA     23   UGUUGU         UUAG   UCAGAAGG   1
    3'->5'               U            UGA        U

The bindings in green highlight positions predicted by PITA
(147-154)

STarMir Site 2

    5'->3'        U      UUUCGUUU    U A      N
    Target   126   ACGAUA         GAUC U GGUCUU     152
                   ||||||         |||| | ||||||
    miRNA     23   UGUUGU         UUAG G UCAGAA     1
    3'->5'               U            U A      GGU
```

PITA predicts 41 binding sites with an overall score -10.25.
There is no binding site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-7	148-154	7:0:0	-16.9	-6.8	-10.09

Position	Structure	Loop Score	ΔG
126-155	3' UGUUGUUUUAGUGAUCAGAAGGU 5' *******:|||***:||||||||* 5' TTCGTTTGATCTTAGGTCTTCCT 3'	25.00	-21.30

RAD51; rs7180135; Mir-197-3p

G-Allele

STarMir predicts 22 binding sites and 2 seed sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
14	707-725	718-724	-28.600	-0.561	-14.289	-14.311	0.380
15	707-721	No	-20.700	-0.076	-14.934	-5.766	0.298
16	707-719	No	-15.100	-0.076	-15.177	0.077	0.338

```
Site 14

    5'->3'        A         CC          A
    Target   707   GCU   UGG  AAGGUGGUGAA   725
                   |||   |||  ||||||||||
    miRNA     22   CGA   ACC  UUCCACCACUU   1
    3'->5'           CCC   UC
The bindings in green here highlight positions predicted by
PITA 717-724 (perfect 8 mer seed match)


Site 15

    5'->3'        A         CC       N
    Target   707   GCU   UGG  AAGGUGG      721
                   |||   |||  |||||||
    miRNA     22   CGA   ACC  UUCCACC      1
    3'->5'           CCC   UC       ACUU


Site 16

    5'->3'        A         CC     N
    Target   707   GCU   UGG  AAGGU        719
                   |||   |||  |||||
    miRNA     22   CGA   ACC  UUCCA        1
    3'->5'           CCC   UC     CCACUU
```

PITA predicts 15 sites with an overall score -12.50.
The following are those covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
G-Allele	hsa-miR-197-3p	716-721	6:1:1	-12.21	-12.94	0.73
G-Allele	hsa-miR-197-3p	717-724	8:0:0	-22	-13.61	-8.38

FindTar predictions

Position	Structure	Loop Score	ΔG
707-725	3' CGACCCACCUCUUCCACCACUU 5' |||***|||**||||||||||* 5' GCT---TGGCCAAGGTGGTGAA 3'	20.00	-25.90

A-Allele

STarMir predicts 21 binding sites and 2 seed site.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
14	707-725	719-724	-22	-1.201	-15.003	-6.997	0.348
15	707-723	No	-18	0	-15.233	-2.767	0.279

```
Site 14

    5'->3'        A          CC   A       A
    Target   707   GCU   UGG  AAG UGGUGAA   725
                   |||   |||  ||| |||||||
    miRNA     22   CGA   ACC  UUC ACCACUU   1
    3'->5'           CCC   UC   C

The bindings in green here highlight positions predicted by
PITA 717-724 (8 mer with a mismatch)

Site 15

    5'->3'        A          CC   A     N
    Target   707   GCU   UGG  AAG UGGUG    723
                   |||   |||  ||| |||||
    miRNA     22   CGA   ACC  UUC ACCAC    1
    3'->5'           CCC   UC   C     UU
```

PITA predicts 14 sites with an overall score -12.48.
There is one binding site covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-197-3p	717-724	8:1:0	-15.4	-13.96	-1.43

Position	Structure	Loop Score	ΔG
707-725	3' CGACCCACCUCUUCCACCACUU 5' |||***|||**||||||* 5' GCT---TGGCCAAGATGGTGAA 3'	20.00	-19.30

ORAI1 ; rs76753792 ; mir-519a-3p

C-Allele

STarMir predicts 4 seed sites and 32 seedless binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
2	69-88	No	-16.700	-4.322	-9.493	-7.207	0.562
3	85-102	No	-18.300	-0.019	-21.657	3.357	0.359

```
Site 2

    5'->3'        A      AC       CAGCC     A
    Target    69   GC CUC      GGA      UGCGC      88
                   || |||      |||      |||||
    miRNA     22   UG GAG      CCU      ACGUG      1
    3'->5'           U   AUUUU              AAA


Site 3

    5'->3'        U             C  G     C
    Target    85   GCGC    AGGGGG UG GCUU    102
                   ||||    |||||| || ||||
    miRNA     22   UGUG    UUUUCC AC UGAA    1
    3'->5'             AGA       U  G     A


The bindings in green highlight positions predicted by PITA
(a 7 mer 96-102 with one mismatch and one G-U wobble)
```

PITA predicts 18 sites with an overall score -10.66.
There is no site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	hsa-miR-519a-3p	96-102	7:1:1	-12.7	-20.64	7.94

T-Allele

STarMir predicts 4 seed sites and 32 seedless binding sites.
The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
2	81-95	No	-15.200	0	-17.396	2.196	0.363
3	81-102	No	-17.900	-0.023	-23.521	5.621	0.381

```
Site 2

    5'->3'        A      UGC   GG  U
    Target    81    GC CUG   AGG  GC       95
                    || |||   |||  ||
    miRNA     22    UG GAU   UCC  CG       1
    3'->5'          UG  A   UU    UA  UGAAA


Site 3

    5'->3'        A     G  C      C  G    C
    Target    81   GC CU UG AGGGGG UG GCUU   102
                   || || || |||||| || ||||
    miRNA     22   UG GA AU UUUCCU AC UGAA   1
    3'->5'           U  G             G    A

The bindings in green highlight positions predicted by PITA (
a 7 mer 96-102 with one mismatch and one G-U wobble)
```

PITA predicts 18 sites with an overall score -10.66.
There is no site covering the SNP position.
The following site may cause the opening of the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
T-Allele	hsa-miR-519a-3p	96-102	7:1:1	-12.8	-22.88	10.08

RAP1 ; rs6573 ; hsa-miR-196a

C-Allele

STarMir predicts 37 seedless binding sites. The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
18	348-370	No	-16.700	-3.971	-5.402	-11.298	0.623

```
STarMir Site 18

    5'->3'        A   UUUUAU     UC  C    U
    Target   348   UCU       CAUGA  CU CCUA   370
                   |||       |||||  || ||||
    miRNA     22   GGG       GUACU  GA GGAU   1
    3'->5'            UUGUU       UU  U
```

The bindings in green highlight positions predicted by PITA
(364-369 a 6 seed match with one mismatch)

PITA predicts 22 binding sites with an overall score -9.02.
There is one binding site covering the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	hsa-miR-196a	364-369	6:1:0	-12.5	-5.03	-7.46

The following site may cause the opening of the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
C-Allele	hsa-miR-196a	368-373	6:1:1	-9.4	-3.22	-6.17

A-Allele

STarMir predicts 37 seedless binding sites and one seed site. The following are those covering the SNP position:

Site ID	Site Position	Seed	ΔG_{hybrid}	ΔG_{nucl}	ΔG_{open}	ΔG_{total}	Site Access
18	348-370	364-369	-21.300	-6.967	-5.051	-16.249	0.616
19	348-368	No	-17.500	-5.900	-4.889	-12.611	0.600

```
STarMir Site 18 seed site

    5'->3'        A   UUUUAU     UC       U
    Target   348   UCU       CAUGA  CUACCUA   370
                   |||       |||||  |||||||
    miRNA     22   GGG       GUACU  GAUGGAU   1
    3'->5'            UUGUU       UU
```

The bindings in green highlight positions predicted by PITA
(364-369 a perfect 6 seed match)

```
STarMir Site 19

    5'->3'        A   UUUUAU     UC     N
    Target   348   UCU       CAUGA  CUACC    368
                   |||       |||||  |||||
    miRNA     22   GGG       GUACU  GAUGG    1
    3'->5'            UUGUU       UU     AU
```

PITA predicts 22 binding sites with an overall score -12.08.
There is one binding site covering the SNP position:

Gene	microRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-196a	364-369	6:0:0	-17.1	-5.04	-12.05

The following site may cause the opening of the SNP position:

Gene	miRNA	Position	Seed	dGduplex	dGopen	ddG
A-Allele	hsa-miR-196a	368-373	6:1:1	-9.4	-3.22	-6.17

FindTar predictions: there is only one binding site predicted for the C-Allele while there are 2 binding sites for the A-Allele; the additional one covers the SNP position:

Position	Structure	Loop Score	ΔG
349-370	3' GGGUUGUUGUACUUUGAUGGAU 5' |::**:|*||||**||||||| 5' CTTTTTATCATGATCCTACCTA 3'	15.00	-21.50

.4 Distribution of metastable conformations and their respective energy values Generated by Luke Day

Contents

1. LIG3 (L = 124)

NM_002311.4 vs. rs4796030
miRNA: miR-221
SNP Pos. 83

	A-allele	C-allele
Target Site:	77 - 98	80 - 98

RNAsubopt and Barrier Setting

Offset	Barrier	A-allele	C-allele	A-allele	C-allele
6.0	1.2	20,646	28,997	349	317

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
A-allele	-20.30	-3.39	-11.77	-8.53
C-allele	-24.60	-0.20	-15.96	-8.64

Local Minima

MFE	A -allele		C-allele	
#Pairings	9		7	
Opening Energy	-10.48		-12.68	
Structure Energy	-33.70		-35.90	
MFE Barrier	6.0		6.00	
Minima				
Less Pairings	28.4%	99	6.3%	20
Equal Pairings	38.5%	134	10.4%	33
Greater Pairings	33.0%	115	83.2%	263
Identical to MFE	4.0%	14	8.5%	27
Minima (-MFE)	**348**		**316**	
Avg. Opening of All Local Minima (excluding MFE)	**-10.38**		**-12.33**	

Less/Equal Pairings than MFE

#Pairings	#Minima A-allele		Target Site Approx. Avg.	#Minima C-allele		Target Site Approx. Avg.
4	0.29%	1	-7.40	1.27%	4	-5.99
5	1.15%	4	-10.49	2.22%	7	-6.98
6	4.02%	14	-11.07	2.85%	9	-10.00
7	8.62%	30	-8.36	**10.44%**	**33**	**-12.41**
8	14.37%	50	-10.01	**24.68%**	**78**	**-13.88**
9	**38.51%**	**134**	**-9.79**	**16.77%**	**53**	**-11.50**
Total	**66.95%**	**233**	**-9.73**	**58.23%**	**184**	**-12.31**
Avg. Barrier	**1.78**			**1.72**		

Local Minima with Less Pairings

	A-allele	C-allele
#LM Less Pairings:	28.4% 99	6.33% 20
Avg. Target Site Energy	-9.65	-8.14
Less Pairings Avg. Barrier Height	1.69	1.74

Local Minima with Less or Equal Approx. Target Site Energy than the MFE

	A -allele	C-allele
Less	61.5% 214	57.9% 183
Equal	4.0% 14	4.4% 14
Less or Equal	65.5% 228	62.3% 197

N+ Test

A-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-30.20	15 **(105)**	2*2.50, 2.30, 2.20, 2*2.10, 2*1.80, 5*1.30, 2*1.20	1.75	-21.08, -16.28, -13.47, -11.96, -10.89, -10.56, -10.49, -10.06, -9.98, 2*-9.76, 2*-8.46, -8.21, -6.98	-11.09
-30.30	14	2*2.60, 2.10, 5*1.80, 1.70, 2*1.60, 3*1.30	1.79	-11.53, -10.76, -7.08, -9.98, 2*-9.86, 2*-10.48, 2*-10.59, -12.03, -6.85, -9.18, -22.36	-10.83
-30.40	2	2.20, 1.80	2.00	-10.36, -9.96	-10.16
-30.50	7	2*2.80, 2.70, 2.40, 2.20, 2.10, 1.80	2.40	-10.36, -10.86, -10.46, -8.51, -9.38, -10.06, -8.76	-9.77
-30.60	8 **(67)**	2*2.90, 2.40, 2.20, 1.80, 1.50, 2*1.30	2.04	-11.06, -10.89, -10.56, -10.48, -10.46, -9.18, -5.85, -7.38	-9.48
-30.70	4	3.00, 1.80, 1.30, 1.60	1.93	-10.56, -10.48, -7.48, -9.18	-9.43
-30.80	4	1.50, 1.30, 1.80, 1.60	1.55	-10.46, -7.58, -10.36, -10.48	-9.72
-30.90	8	2*3.20, 2.2, 1.90, 1.80, 2*1.30, 1.40	2.04	3*-7.68, -9.38, -10.86, -11.33, -10.46, -9.78	-9.36
-31.00	3	1.40, 3.30, 1.80	2.17	-7.78, -10.86, -10.56	-9.73
-31.10	5	2*2.30, 1.80, 1.60, 1.30	1.86	-11.53, -9.98, -9.18, 2*-7.88	-9.29
-31.20	4	1.31, 2.30, 3.50, 1.60	2.18	-12.43, -7.98, -14.47, -9.46	-11.09
-31.30	4	2*1.80, 1.30, 2.60	1.43	-10.86, -9.38, -9.78, -8.08	-9.53
-31.40	2	2.60, 1.90	2.25	2*-8.18	-8.18
-31.50	4	2*1.30, 2.10, 2.30	1.43	-10.49, -11.96, -9.98, -9.76	-10.55
-31.60	6 **(21)**	3.70, 3.90, 2*1.30, 2.00, 3.20	2.57	-12.03, -6.85, -10.48, -9.86, 2*-10.59	-10.06
-31.70	2	2.30, 1.30	1.80	-9.96, -9.78	-9.87
-31.80	1	2.30	2.30	-10.06	-10.06

-31.90	2	2.10, 1.80	1.95	-10.89, -9.98	-10.44
-32.00	1 **(10)**	4.30	4.30	-10.48	-10.48
-32.10	1	2.20	2.20	-10.36	-10.36
-32.20	1	2.40	2.40	-10.46	-10.46
-32.30	1	3.20	3.20	-10.56	-10.56
-32.40	2	1.80, 1.60	1.70	-10.48, -9.18	-9.83
-32.60	2	4.70, 3.70	4.20	-10.86, -9.38	-10.12
-33.00	1	1.30	1.30	-9.78	-9.78
-33.20	1	2.30	2.30	-9.98	-9.98
MFE: -33.70	1	6.00	6.00	-10.48	-10.48
-3283.9 Avg. -30.98	**106**	**220.81**	**2.08**	**-1073.4**	**-10.13**

A-allele - N+ Averages

Test #LM-MFE	Avg. Structure Energy	Avg. Barrier	Avg. Opening Energy
10	-324.8/10 -32.48	27.5/10 2.75	-101.52/10 10.15
21	-673.4/21 -32.07	52.7/21 2.51	-212.59/21 -10.12
67	-2098.7/67 -31.32	142.71/67 2.13	-656.52/67 -9.80
105	-3250.2/105 -30.95	214.81/105 2.04	-1063.26/105 -10.13

N+ Test

C-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-32.40	9 **(107)**	2* 2.50, 2*2.40, 2*1.80, 1.60, 2*1.20	1.83	-16.20, 2*-15.30, -14.43, -12.76, -12.36, -10.88, -10.48, -6.30	-12.67
-32.50	7	2*2.60, 2.40, 2.10, 1.80, 1.60, 1.20	2.04	-15.30, -14.53, -12.46, 2*-12.68, -12.38, -12.36	-13.20
-32.60	9	3* 2.70, 2.40, 3*2.10, 2.00, 1.80	2.29	-15.87, -13.29, -12.46, -12.38, -11.59, -11.20, -10.86, -10.61, -6.80	-11.67
-32.70	8	2.80, 2.40, 2.30, 2.10, 2*1.60, 2*1.20	1.90	-20.23, -16.20, -13.16, -12.38, -11.58, -10.96, -9.48, -7.95	-12.37
-32.80	8	2*2.90, 2.30, 1.90, 2*1.80, 1.50, 1.40	2.06	-12.76, -12.68, -12.36, -11.68, -11.20, -10.88, 2*-9.58	-11.34
-32.90	8 **(66)**	2* 3.00, 2.70, 3*1.80, 1.30, 1.20	2.08	-16.23, -13.50, -12.76, -12.68, -12.46, -11.78, -11.16, -9.68	-12.53
-33.00	7	2*3.10, 2.10, 1.80, 1.60, 2* 1.30	2.04	-13.60, -13.29, -12.68, -11.26, 2*-9.78, -8.25	-11.23
-33.10	1	1.60	1.60	-11.58	-11.58
-33.20	5	3.20, 1.80, 1.40,	1.78	-16.50, -15.46, -15.30, -12.76, -11.68	-14.34

		1.30, 1.20			
-33.30	6	3.40, 2.20, 1.90, 1.80, 1.40, 1.20	1.98	-16.23, -13.50, -11.78, 3*-10.08	-11.96
-33.40	4	2.10, 1.80, 1.30, 1.20	1.60	-18.16, -17.13, -16.23, -13.60	-16.28
-33.50	4	3.60, 3.20, 1.80, 1.60	2.55	-13.93, -12.38, -11.58, -10.28	-12.04
-33.60	3	3.70, 1.60, 1.40	2.23	-15.30, -11.86, -11.68	-12.95
-33.70	6	3.80, 2.60, 2*1.80, 2*1.20	1.56	-16.23, -15.30, -14.43, -13.50, -11.78, -10.48	-13.62
-33.80	3 **(22)**	3.20, 1.80, 1.20	2.07	-14.53, -13.60, -12.68	-13.60
-33.90	1	4.00	4.00	-12.38	-12.38
-34.00	2	1.80, 1.20	1.50	-16.20, -15.30	-15.75
-34.10	2	2*2.40	2.40	-12.36, -10.88	-11.62
-34.20	2	4.30, 3.20	3.75	-12.68, -12.46	-12.57
-34.30	2	2.10, 1.80	1.95	-13.29, -12.38	-12.84
-34.50	1 **(10)**	4.60	4.60	-12.76	-12.76
-34.60	1	1.80	1.80	-12.68	-12.68
-34.80	1	1.60	1.60	-11.58	-11.58
-34.90	1	1.40	1.40	-11.68	-11.68
-35.00	3	4.20, 1.80, 1.20	2.40	-16.23, -13.50, -11.78	-13.84
-35.10	1	2.10	2.10	-13.60	-13.60
-35.30	1	5.40	5.40	-15.30	-15.30
-35.60	1	5.20	5.20	-12.38	-12.38
MFE: -35.90	1	6.00	6.00	-12.68	-12.68
-3594.4 Avg. -33.28	**108**	**237.7**	**2.20**	**1374.8**	**-12.73**

C-allele - N+ Averages

Test #LM-MFE	**Avg. Structure Energy**	**Avg. Barrier**	**Avg. Opening Energy**
10	-349.8/10 -34.98	29.3/10 2.93	131.49/10 -13.15
22	-758.3/22 -34.47	58.7/22 2.67	-290.23/22 -13.19
66	-2222/66 -33.67	147.7/66 2.24	-858.00/66 -13.00
107	-3558.5/107 -33.26	231.7/107 2.17	-1362.12/107 -12.73

Ordered by Deepest Local Minima

A-allele

Barrier	#LM	Opening	Avg. Opening
4.70	1	-10.86	-10.86
4.30	1	-10.48	-10.48
3.90	1	-6.85	-6.85
3.70	2	-12.03, -9.38	-10.71
3.50	1	-14.47	-14.47
3.30	1	-10.86	-10.86
3.20	4 **(11)**	-9.38, -10.48, -10.56, -10.86	-10.32
3.00	1	-10.56	-10.56
2.90	2	-10.56, -10.48	-10.52
2.80	2	-9.38, -10.86	-10.12
2.70	1	-8.76	-8.76
2.60	4 **(21)**	-22.36, -8.08, -8.18, -11.53	-12.54
2.50	2	-21.08, -10.56	-15.82
2.40	5	4*-10.46, -5.35	-9.44
2.30	14	-7.98, -8.26, 3*-9.96, 4*-9.98, 3*-10.06, -11.53, -20.88	-10.62
2.20	9	-5.85, -6.85, -7.68, 4*-10.36, -12.03, -13.47	-9.70
2.10	16 **(67)**	-5.05, -6.68, -8.21, -8.51, -9.98, -9.09, 3*-10.89, -10.48, 3*-11.96, -15.31, -15.61, -22.36	-11.24
2.00	7	-8.18, -8.26, -10.06, 2*-10.59, -16.66, -21.08	-10.69
1.90	8	-7.68, -8.08, 2*-8.18, -9.96, -10.86, -11.69, -15.98	-10.08
1.80	42 **(124)**	2*-6.85, 2*-7.68, -7.76, 2*-7.88, 2*-7.98, -8.08, 2*-8.18, 3*-9.38, -9.96, 3*-9.98, -10.06, -10.36, -10.46, 3*-10.48, -10.56, 3*-10.59, -10.76, 4*-10.86, -10.89, -11.53, -11.96, 2*-12.03, 2*-14.47, -20.88	-10.19

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
11	39.9/11 3.63	-116.21/11 -10.56
21	67.4/21 3.21	-226.96/21 -10.81
67	170.0/67 2.54	-721.57/67 -10.77
124	274.8/124 2.22	-1315.73/124 -10.61

C-allele

Barrier	#LM	Opening	Avg. Opening
5.40	1	-15.30	-15.30
5.20	1	-12.38	-12.38
4.60	1	-12.76	-12.76
4.30	1	-12.68	-12.68
4.20	1	-11.78	-11.78
4.00	1	-12.38	-12.38
3.80	1	-15.30	-15.30
3.70	1	-15.30	-15.30
3.60	1	-13.93	-13.93
3.40	1 **(10)**	-11.78	-11.78
3.20	4	-12.38, -12.46, -12.68, -15.30	-13.21
3.10	2	-8.25, -11.26	-9.76
3.00	2	-11.78, -12.76	-12.27
2.90	1	-12.68	-12.68
2.80	1 **(20)**	-20.23	-20.23
2.70	4	-6.80, -11.16, -12.46, -15.87	-11.57
2.60	3	-10.48, -12.38, -12.46	-11.77
2.50	2	-6.30, -12.76	-9.53
2.40	10	-6.86, -9.08, -10.86, 2*-10.88, -10.96, 3*-12.36, -15.30	-11.19
2.30	4	-9.48, -11.20, -15.30, -20.23	-14.05
2.20	4	-10.08, -10.73, -12.46, -15.13	-12.10
2.10	18 **(65)**	-7.20, -10.48, -10.61, -10.88, -11.59, -12.36, -12.38, -12.43, -12.68, -13.16, 3*-13.29, 4*-13.60, -17.71	-12.54
2.00	4	-10.16, -11.20, -11.78, -15.30	-12.11
1.90	10	-6.00, 4*-8.58, -8.60, -9.33, -9.58, -10.08, -13.93	-9.18
1.80	43	-6.95, -8.25, -8.88, -9.96, 2*-10.08, 3*-10.28, -10.48, -10.88, -11.26, 2*-11.78, -12.36, 3*-12.38, -12.46, 3*-12.68, -12.76, -13.29, 4*-13.60, -13.93, 4*-15.30, 9*-13.23, -20.23	-12.47

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	42.2/10 4.22	-133.59/10 -13.36
20	73/20 3.65	-263.37/20 -13.17
65	177.1/65 2.72	-811.86/65 -12.49
122	281.8/122 2.31	-1509.76/122 -12.38

Ordered by Opening Energy

A-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-4.15	1	1.20	1.20
-4.35	1	1.30	1.30
-4.55	1	1.60	1.60
-5.05	1	2.10	2.10
-5.35	2	2.40, 1.40	1.90
-5.68	1	1.20	1.20
-5.85	3 **(10)**	2.20, 1.60, 1.20	1.67
-5.98	3	1.30, 1.40, 1.50	1.40
-6.18	1	1.70	1.70
-6.38	1	1.20	1.20
-6.48	3	1.50, 1.30, 1.20	1.33
-6.58	1	1.70	1.70
-6.68	2 **(21)**	2.10, 1.50	1.80
-6.85	4	3.90, 2.20, 2*1.80	2.43
-6.88	1	1.50	1.50
-6.98	1	1.20	1.20
-7.01	1	1.30	1.30
-7.08	3	1.70, 1.30, 1.20	1.40
-7.16	1	1.20	1.20
-7.26	1	1.30	1.30
-7.38	3	2*1.30, 1.20	1.27
-7.40	1	1.30	1.30
-7.46	2	1.30	1.30
-7.48	3	3*1.30	1.30
-7.56	3	1.60, 1.50, 1.30	1.47
-7.58	3	3*1.30	1.30
-7.66	1	1.70	1.70
-7.68	10	2.20, 1.90, 2*1.80, 2*1.50, 3*1.40, 1.30	1.62
-7.76	1 **(60)**	1.80	1.80
-7.78	4	3*1.40, 1.20	1.35
-7.86	2	1.30, 1.20	1.25
-7.88	8	2*1.80, 1.70, 5*1.30	1.48
-7.91	1	1.60	1.60
-7.96	1	1.30	1.30
-7.98	4	2.30, 2*1.80, 1.40	1.83
-8.06	1	1.30	1.30
-8.08	4	2.60, 1.90, 1.80, 1.50	1.95
-8.16	1	1.40	1.40
-8.18	8	2.60, 2.00, 2*1.90, 2*1.80, 2*1.60	1.90
-8.21	2	2.10, 1.20	1.65
-8.26	2	2.30, 2.00	2.15
-8.33	2 **(100)**	1.40, 1.30	1.35

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	16.2/10 1.62	-52.03/10 -5.20
21	32.6/21 1.55	-121.91/21 -5.81
60	92.6/60 1.54	-410.38/60 -6.84
100	158.5/100 1.59	-731.63/100 -7.32

C-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-5.50	1	1.70	1.70
-5.66	1	1.20	1.20
-6.00	1	1.90	1.90
-6.06	1	1.60	1.60
-6.30	2	2.50, 1.20	1.85
-6.35	2	2*1.20	1.20
-6.40	1	1.30	1.30
-6.45	2 **(11)**	2*1.30	1.30
-6.80	2	2.70, 1.40	2.05
-6.86	1	2.40	2.40
-6.95	1	1.80	1.80
-7.20	1	2.10	2.10
-7.88	1	1.20	1.20
-7.95	1	1.20	1.20
-7.98	1	1.30	1.30
-8.08	1 **(20)**	1.30	1.30
-8.25	3	3.10, 1.80, 1.40	2.10
-8.28	1	1.30	1.30
-8.58	4	4*1.90	1.90
-8.60	1	1.90	1.90
-8.68	1	1.50	1.50
-8.88	1	1.80	1.80
-8.93	1	1.40	1.40
-8.98	1	1.70	1.70
-9.08	1	2.40	2.40
-9.30	1	1.70	1.70
-9.33	1	1.90	1.90
-9.48	2	2.30, 1.50	1.90
-9.56	3	3*1.40	1.40
-9.58	6	1.90, 1.60, 2*1.50, 2*1.20	2.23
-9.60	1	1.20	1.20
-9.66	1	1.50	1.50
-9.68	3	3*1.30	1.30

-9.76	2	1.60, 1.30	1.45
-9.78	6 **(60)**	6*1.30	1.30
-9.86	2	2*1.30	1.30
-9.96	2	1.80, 1.40	1.60
-10.08	12	2.20, 1.90, 2*1.80, 2*1.70, 3*1.40, 3*1.30	1.60
-10.13	1	1.20	1.20
-10.16	1	2.00	2.00
-10.23	1	1.30	1.30
-10.28	4	3*1.80, 1.50	1.73
-10.48	4	2.60, 2.10, 1.80, 1.70	2.05
-10.61	2	2.10, 1.40	1.75
-10.73	1	2.20	2.20
-10.86	2	2.40, 1.40	1.90
-10.88	6	2*2.40, 2.10, 1.80, 1.30, 1.20	1.87
-10.96	3 **(101)**	2.40, 1.50, 1.20	1.70

C-allele - Averages

Test #LM-MFE	**Avg. Barrier**	**Avg. Opening Energy**
11	16.4/11 1.49	-67.82/11 -6.17
20	31.8/20 1.59	-134.32/20 -6.72
60	95.5/60 1.59	-505.07/60 -8.42
101	165.9/101 1.64	-931.06/101 -9.22

2. CBR1 (L = 284)

NM_001757.2 vs. rs9024
miRNA: miR-574-5p
SNP Pos. 133
Target site: 121 - 162

RNAsubopt and Barrier Setting

Offset	Barrier	G-allele	A-allele	G-allele	A-allele
6.0	1.4	10,987,436	16,209,366	7,457	11,187

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_disrupt	dG_total
G-allele	-27.80	-5.36	-8.09	-19.72
A-allele	-27.80	-5.22	-5.96	-21.84

Local Minima

MFE	**G-allele**	**A-allele**
#Pairings	14	12
Opening Energy	-10.90	-9.90
Structure Energy	-52.10	-51.10
MFE Barrier	6.00	6.00
Minima		
Less Pairings	39.6% 2,954	20.3% 2,274
Equal Pairings	18.6% 1,385	30.7% 3,431
Greater Pairings	41.8% 3,117	49.0% 5,481
Identical to MFE	12.8% 951	8.5% 951
Minima (- MFE)	**7,456**	**11,186**
Approx. Target Energy Avg. (excluding MFE)	**-10.58**	**-9.27**

Less/Equal Pairings than MFE Target Site

#Pairings	#Minima G-Allele	Target Site Approx. Avg.	#Minima A-allele	Target Site Approx. Avg.
4	-	-	0.07% 8	-2.20
6	0.28% 21	-4.39	0.61% 68	-4.62
7	0.25% 19	-4.57	1.25% 140	-4.48
8	0.15% 11	-5.86	3.22% 360	-5.68
9	0.52% 39	-5.23	0.75% 84	-5.81
10	1.44% 107	-5.64	8.31% 929	-7.10
11	1.22% 91	-6.48	6.12% 685	-7.36
12	24.75% 1,845	-7.64	**30.67% 3,431**	**-8.25**
13	11.01% 821	-8.38	**14.44% 1,615**	**-8.55**
14	**18.58% 1,385**	**-10.68**	**3.41% 382**	**-9.00**
Total Minima	**58.19% 4,339**	**-8.62**	**68.85% 7,702**	**-7.88**
Avg. Barrier	**1.83**		**1.82**	

Local Minima with Less Pairings

	G-allele	A-allele
#LM Less Pairings:	36.6% 2,954	20.3% 2,274
Avg. Target Site Energy	-7.66	-6.65
Less Pairings Avg. Barrier Height	1.82	1.82

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	G-allele	A-allele
Less	56.37% 4,203	58.04% 6,492
Equal	13.37% 997	8.93% 999
Less or Equal	69.74% 5,200	66.97% 7,491

N+ Test

G-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-49.80	25 **(123)**	10*3.70, 2*3.20, 2.10, 2.30, 2*2.40, 7*2.20, 2*1.70	2.86	5*-18.70, 5*-12.75, -11.10, 6*-10.90, 5*-8.20, 2*-7.30, -7.00	-11.85
-49.90	23	2*3.80, 3.70, 3.40, 2*3.20, 2.90, 4*2.20, 2*2.40, 3*2.10, 2.00, 2*1.90, 1.80, 2*1.70, 1.60	2.46	5*-12.75, -12.00, -11.10, 5*-10.90, -9.20, 5*-8.20, 2*7.30, 3*-7.00	-9.88
-50.00	15	3*3.90, 3.80, 3.40, 3.20, 2.40, 2*2.30, 2.10, 2*2.20, 3*1.80	3.77	2*-18.70, 2*-12.75, 4*-10.90, 9.80, 2*-8.20, 2*-7.60, -7.30, -7.00	-10.81
-50.10	13 **(60)**	5*4.00, 3.90, 3.80, 2*3.20, 3*2.10, 1.90	3.25	2*-12.75, -11.60, -11.10, 3*-8.20, -7.30, 3*-7.00, 2*-5.80	-8.67
-50.20	9	2*4.10, 3.70, 2*2.90, 3*2.40, 1.80	2.97	2*-18.70, -12.75, -10.90, -9.00, -8.20, 3*-7.30	-11.13
-50.30	6	4.20, 3.90, 3.80, 3.70, 2.90, 2.10	3.43	2*-18.70, -12.75, -10.90, -8.20, -7.30	-12.76
-50.40	10	4*4.30, 2*3.80, 3.40, 2*2.90, 2.10	3.61	3*-12.75, -9.20, 3*-8.20, 3*-7.30	-9.40
-50.50	6 **(22)**	2*4.40, 3.80, 2.90, 2.30, 1.80	3.27	2*-18.70, -12.75, -10.90, -8.20, -7.60	-12.81
-50.60	8 **(16)**	4*4.50, 2*3.80, 2*2.10	3.73	3*-12.75, 3*-8.20, -10.90, -5.80	-9.94
-50.70	1	1.90	1.90	-10.90	-10.90
-50.80	4	4.70, 3.90, 2.30,	3.25	4*-10.90	-10.90

		2.10			
-51.30	2	3.90, 2.30	3.10	2*-10.90	-10.90
-51.60	1	5.50	5.50	-10.90	-10.90
MFE: -52.10	1	6.00	6.00	-10.90	-10.90
-6219.6 Avg. -50.16	**124**	**376.6**	**3.04**	**1323.55**	**-10.67**

G-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
16	-812.9/18 -50.80	56.4/16 3.53	166.75/16 -10.42
60	-3024.8/60 -50.41	201.4/60 3.36	-626.95/60 -10.45
22	-1115.9/22 -50.72	76/22 3.45	-243.6/22 -11.07
123	-6167.5/123 -50.14	370.6/123 3.01	-1312.65/123 -10.67

A-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-49.20	20 **(113)**	4.10, 3.70, 2*2.90, 3*2.40, 2.20, 3*2.00, 5*1.90, 4*1.80	2.29	2*-17.70, -11.75, 3*-10.42, -9.90, -8.00, 2*-7.60, 3*-7.30, -7.20, 4*-6.30, 2*-5.80	-8.87
-49.30	19	2*4.20, 3.90, 2*3.80, 3.70, 2.90, 2*2.80, 2*2.20, 2*2.10, 2*2.00, 1.90, 3*1.80	2.74	2*-17.70, -11.75, -10.42, -9.90, 2*7.60, 7*-7.30, -7.20, -6.30, 2*-5.80, -4.50	-8.60
-49.40	16 **(74)**	3*4.30, 2*3.80, 3.40, 2*2.90, 2.20, 3*2.00, 3*1.90, 1.80	2.84	3*-11.75, 3*-10.42, -8.20, -7.60, 3*7.20, 3*6.30, 2*5.80	-8.40
-49.50	11	3*4.40, 3.80, 2*3.10, 2.90, 2.30, 2.20, 2.10, 1.80	3.15	2*-17.70, -11.75, -10.55, -9.90, -7.60, 3*-7.30, -7.20, -5.80	-10.00
-49.60	14	3*4.50, 4.30, 3.90, 2*3.80, 3.40, 3*2.20, 2*2.10, 1.80	3.24	3*11.75, 9.90, -7.60, 5*-7.30, 3*7.20, -5.80	-8.33
-49.70	7	2.90, 2.40, 2*2.20, 2.10, 1.90, 1.70	2.20	-9.90, 5*-7.30, -5.80	-7.46
-49.80	6	4.70, 2*3.90, 2.30, 2.20, 2.10	3.18	4*-9.90, 2*-7.30	-9.03
-49.90	2 **(20)**	4.80, 2.10	3.45	2*-7.30	-7.30

-50.00	3	2.40, 2*1.80	2.00	2*-7.60, -7.30	-7.50
-50.10	3	2*5.00, 3.90	4.63	-7.30, 2*-5.80	-6.30
-50.20	3 **(12)**	2*5.10, 2.90	4.36	3*-7.30	-7.30
-50.30	3	5.20, 3.90, 2.30	3.80	2*-9.90, -7.30	-9.03
-50.40	3	2*5.30, 2.10	4.23	3*-7.30	-7.30
-50.50	1	1.80	1.80	-7.60	-7.60
-50.60	2	2*5.50	5.50	-9.90, -5.80	-7.85
MFE: -51.10	1	6.00	6.00	-9.90	-9.90
-5652.3 Avg. -49.58	**114**	**340.2**	**2.98**	**-968.44**	**-8.57**

A-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
12	-604.4/12 -50.37	50/12 4.17	94.2/12 -7.85
20	-1004.5/20 -50.23	76.8/20 3.84	-150.2/20 -7.51
74	-3680.5/74 -49.74	236.5/74 3.20	-617.76/74 -8.35
113	-5601.2/113 -49.57	334.2/113 2.96	-958.54/113 -8.48

Ordered by Deepest Local Minima

G-allele

Barrier	#LM	Opening	Avg. Opening
5.50	1	-10.90	-10.90
4.70	1	-10.90	-10.90
4.50	4	-5.80, -10.90, 2*-12.75,	-10.55
4.40	2	-18.70, -12.75	-15.73
4.30	4 **(12)**	2*-7.30, 2*-12.75	-10.02
4.20	1	-7.30	-7.30
4.10	2	2*-7.30	-7.30
4.00	5 **(20)**	2*-5.80, -11.10, -11.60, -12.75	-9.41
3.90	7	-7.30, 3*-10.90, 2*-12.75, -18.70	-12.03
3.80	10	-7.30, 8*-8.20, -10.90	-8.38
3.70	13	-7.30, 2*-8.20, 3*-10.90, 2*-11.10, 2*-12.75, 3*-18.70	-12.32
3.60	4	4*-10.90	-10.90
3.50	12 **(66)**	2*-7.30, 7*-10.90, -11.50, -11.60, -18.70	-11.06
3.40	14	-5.80, -8.20, 3*-9.20, 2*-9.80, -10.00, -10.55, 2*-10.90, -12.75, 2*-18.70	-10.98
3.30	12	-5.80, -7.90, -8.20, 2*-9.93, -10.20, 2*-10.90, -11.10, -11.30, -12.75, -18.70	-10.63
3.20	36 **(128)**	-4.50, -5.80, 10*-7.00, -7.30, 5*-8.20, -9.10, -9.93, -10.00, 2*-10.90, -11.60, 5*-12.75, 7*-18.70	-10.71

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
12	54.2/12 4.52	-135.55/12 -11.30
20	86.6/20 4.33	-204.5/20 -10.23
66	256.4/66 3.88	-709/66 -10.74
128	458.80/128 3.58	-1375.99/128 -10.75

A-allele

Barrier	#LM	Opening	Avg. Opening
5.50	2	-5.80, -9.90	-7.85
5.30	2	2*-7.30	-7.30
5.20	1	-7.30	-7.30
5.10	2	2*-7.30	-7.30
5.00	2	2*5.80	-5.80
4.80	1 **(10)**	-7.30	-7.30
4.70	1	-9.90	-9.90
4.50	3	-9.90, 2*-11.75	-11.13
4.40	3	-10.55, -11.75, -17.70	-13.33
4.30	4 **(21)**	-5.80, -7.30, 2*-11.75	-9.15
4.20	2	-5.80, -7.30	-6.55
4.10	1	-7.30	-7.30
4.00	8	-5.80, 4*-7.30, -10.10, -10.60, -11.75	-8.43
3.90	16	7*-7.30, -8.40, 3*-9.90, 2*-10.55, 2*-11.75, -17.70	-9.47
3.80	15 **(63)**	-4.50, 2*-5.80, 8*-7.20, 3*-7.30, -9.90	-7.03
3.70	19	2*-4.50, 2*-7.20, 4*-7.30, -8.40, 3*-9.90, 2*-10.10, 2*-11.75, 3*-17.70	-9.87
3.60	11	-4.90, -5.80, 4*-7.30, -8.40, 4*-9.90	-7.99
3.50	13 **(106)**	-4.50, 2*-7.30, 7*-9.90, -10.50, -10.60, -17.70	-9.78

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	51.8/10 5.18	-71.1/10 -7.11
21	100.4/21 4.78	-191/21 -9.10
63	264.3/63 4.20	-535.85/63 -8.51
106	419.7/106 3.96	-938.45/106 -8.85

Ordered by Opening Energy

G-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-2.20	1	1.60	1.60
-2.90	1	1.60	1.60
-3.40	4	3*1.50, 1.40	1.48
-3.50	4 **(10)**	1.90, 2*1.70, 1.50	1.70
-3.90	10 **(20)**	4*1.70, 3*1.60, 1.50, 2*1.40	1.59
-4.00	9	2*1.70, 1.50, 6*1.40	1.48
-4.10	12	2.80, 2*2.30, 1.90, 1.80, 1.70, 2*1.60, 2*1.50, 2*1.40	1.82
-4.50	26 **(67)**	3.20, 2*2.70, 2.50, 2.30, 2.20, 2.10, 4*2.00, 1.90, 3*1.80, 2*1.70, 2*1.60, 3*1.50, 4*1.40	1.91
-4.70	6	2.40, 2*1.90, 2*1.50, 1.40	1.77
-4.90	8	2.60, 2*2.10, 1.70, 1.60, 1.50, 2*1.40	1.80
-5.00	10	2.70, 2*2.20, 1.80, 1.70, 1.60, 2*1.50, 2*1.40	1.80
-5.20	3	3*1.50	1.50
-5.30	18 **(112)**	1.80, 8*1.60, 2*1.50, 7*1.40	1.52

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	15.9/10 1.59	-32.7/10 -3.27
20	31.8/20 1.59	-71.7/20 -3.59
67	116.6/67 1.74	-273.9/67 -4.09
112	191.5/112 1.71	-502.3/112 -4.48

A-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-2.20	8	5*1.60, 1.50, 2*1.40	1.36
-2.40	1	1.40	1.40
-2.60	1 **(10)**	1.60	1.60
-2.90	17 **(27)**	3*1.90, 5*1.70, 3*1.60, 2*1.50, 4*1.40	1.62
-3.00	6	6*1.40	1.40
-3.40	8	2.40, 2*1.90, 1.50, 4*1.40	1.66
-3.50	26 **(67)**	4*1.40, 3*1.50, 2*1.60, 2*1.70, 3*1.80, 12*1.90	1.73
-3.60	1	1.40	1.40
-3.70	1	1.50	1.50

-3.90	16	2.60, 2*2.10, 3*1.90, 2*1.70, 2*1.60, 2*1.50, 4*1.40	1.73
-4.00	26 **(111)**	5*1.40, 5*1.50, 6*1.60, 2*1.70, 2*1.80, 3*1.90, 2*2.20, 2.70	1.69

A-allele - Averages

Test #LM-MFE	**Avg. Barrier**	**Avg. Opening Energy**
10	15.3/10 1.53	-22.6/10 -2.26
27	42.9/27 1.59	-71.9/27 -2.66
67	109.5/67 1.63	-208.1/67 -3.11
111	184/111 1.66	-381.8/111 -3.44

3. HTR3E (L = 302)

NM_001256614.1 vs. rs56109847
miRNA: miR-510-5p
SNP Pos. 76

	G-allele	A-allele
Target Site:	50 - 80	50 - 80

RNAsubopt and Barrier Setting

Offset	Barrier	G-allele	A-allele	G-allele	A-allele
6.0	1.4	260,869	239,418	1,174	997

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_disrupt	dG_total
G-allele	-28.70	-5.505	-12.645	-16.055
A-allele	-22.50	-2.464	-8.261	-14.239

Local Minima

MFE	G -allele		A-allele	
#Pairings	15		9	
Opening Energy	-24.47		-10.96	
Structure Energy	-66.90		-63.90	
MFE Barrier	6.00		6.00	
Minima				
Less Pairings	10.91%	128	7.53%	75
Equal Pairings	78.86%	925	24.29%	242
Greater Pairings	10.23%	120	68.17%	679
Target Site Identical to MFE	13.73%	161	13.05%	130
Minima (- MFE)	**1,173**		**996**	
Avg. Opening of All Local Minima (excluding MFE)	**-21.74**		**-10.56**	

Less/Equal Pairings than MFE

#Pairings	#Minima G-allele		Target Site Approx. Avg.	#Minima A-allele		Target Site Approx. Avg.
6	0.09%	1	-5.40	0.10%	1	-2.20
7	-		-	0.50%	5	-5.04
8	1.02%	12	-9.42	6.93%	69	-6.03
9	0.34%	4	-10.58	24.3%	242	-9.43
10	0.68%	8	-7.90	**22.5%**	**224**	**-6.67**
11	1.71%	20	-10.50	**4.82%**	**48**	**-7.44**
12	2.30%	27	-9.68	**7.13%**	**71**	**-8.73**
13	1.36%	16	-17.29	**9.84%**	**98**	**-10.88**
14	3.41%	40	-21.63	**20.38%**	**203**	**-18.10**
Total:	**10.91%**	**128**	**-14.35**	**96.5%**	**961**	**-10.34**
Avg. Barrier:	**1.79**			**1.79**		

Local Minima with Less Pairings

	G-allele	A-allele
#LM Less Pairings:	10.91% 128	7.53% 75
Avg. Target Site Energy	-14.35	-5.91

Local Minima with Less or Equal Approx. Target Site Energy than the MFE

	G-allele	A-allele
Less	36.15% 424	595
Equal	27.54% 323	130
Less or Equal	63.68% 747	725

N+ Test

G-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-64.40	12 **(108)**	4*2.10, 4*1.70, 4*1.50	1.77	4*-25.40, 8*-21.37	-22.71
-64.50	14	4*2.10, 2*1.90, 4*1.60, 4*1.50	1.76	8*-21.47, 6*-21.37	-21.43
-64.60	10	4*2.10, 4*1.60, 2*1.90	1.86	6*-21.47, 4*-21.37	-21.43
-64.70	4	4*2.10	2.10	4*-21.47	-21.47
-64.80	4	4*2.30	2.30	4*-21.37	-21.37
-64.90	10 **(64)**	4*2.90, 6*2.30	2.54	4*-21.47, 6*-21.37	-21.41
-65.00	6	6*2.60	2.60	6*-21.47	-21.47
-65.10	8	4*2.30, 4*1.70	2.00	4*-25.30, 4*-21.37	-23.44
-65.20	8	4*2.60, 4*1.70	2.15	4*-25.40, 4*-21.47	-23.44
-65.30	4	4*2.30	2.30	4*-21.37	-21.37
-65.40	4	4*2.50	2.50	4*-21.47	-21.47
-65.80	4	4*1.70	1.70	4*-25.30	-25.30
-65.90	4 **(20)**	4*1.70	1.70	4*-25.40	-25.40
-66.10	4	4*2.10	2.10	4*-21.37	-21.37
-66.20	4 **(12)**	4*2.10	2.10	4*-21.47	-21.47
-66.80	4	4*2.30	2.30	4*-21.37	-21.37
MFE: -66.90	4	6.00, 4.70, 2*2.90	4.13	4*-21.47	-21.47
Total: -7044.8 Avg. -65.23	**108**	**1.99**		**-22.15**	

G-allele - N+ Averages

Test #LM-MFE	Avg. Structure Energy	Avg. Barrier	Avg. Opening Energy
12	-799.6/12 -66.63	34.1/12 2.84	-257.24/12 -21.44
20	-1327.6/20 -66.38	49.3/20 2.47	-444.32/20 -22.22
64	-4205/64 -65.70	149.5/64 2.34	-1433.96/64 -22.41
108	-7044.8/108 -65.23	231.5/108 2.14	-2392.16/108 -22.15

N+ Test

A-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-61.00	19 **(118)**	2*3.10, 6*2.60, 2.50, 2.30, 2*2.20, 2*2.00, 1.60, 4*1.40	2.22	6*-17.47, 3*-10.96, 2*-10.00, 2*-8.06, -7.90, 3*-7.20, -6.20, -5.80	-11.33
-61.10	9	2.10, 2*1.90, 2*1.80, 2*1.70, 2*1.60	1.79	2*-13.40, 3*-10.96, -8.86, 3*-7.20	-10.02
-61.20	22	5*3.30, 2.80, 4*2.60, 2*2.10, 2*1.80, 4*1.70, 4*1.40	2.27	4*-21.40, 4*-17.47, 4*-10.96, 2*-7.90, 4*-7.20, 2*-6.90, -6.20, -5.50	-12.25
-61.30	6	2*3.40, 2.80, 2.60 2*2.10	2.73	-10.96, 2*-10.40, -9.00, -7.50, -7.20	-9.24
-61.40	10 **(62)**	4*2.50, 3*2.10, 1.50, 2*1.40	2.06	4*-17.47, 2*-10.40, -8.30, -6.20, 2*-5.40	-11.60
-61.50	6	2*3.60, 2*2.10, 2*1.50	2.40	3*-10.96, 3*-7.20	-9.08
-61.60	5	3.70, 2.70, 2.10, 2*1.60	2.34	-10.96, -8.66, 2*-8.30, -7.20	-8.68
-61.70	6	3*2.10, 3*1.40	1.75	-10.96, 2*-10.00, -8.10, -7.90, -7.20	-9.03
-61.80	2	2.30, 2.10	2.20	-8.86, -5.80	-7.33
-61.90	6	4*1.70, 2*2.90	2.10	4*-21.40, -10.96, -7.20	-17.29
-62.00	3	2*2.60, 1.40	2.20	-10.96, 2*-7.70	-8.79
-62.10	4	2*4.20, 2.50, 1.50	3.10	2*-10.40, -8.30, -5.40	-8.63
-62.20	7 **(20)**	2*2.60, 4*2.10, 1.40	2.14	4*-17.47, -10.96, -7.20, -6.20	-13.46
-62.40	3	4.50, 2*2.50	3.17	-10.96, -8.10, -7.20	-8.75
-62.50	1 **(10)**	4.60	4.60	-5.80	-5.80
-62.90	5	5.00, 4.70, 2*2.90, 1.40	3.38	4*-17.47, -6.20	-15.22
-63.20	2	2*2.10	2.10	-10.96, -7.20	-9.08
MFE: -63.90	2	6.00, 5.70	5.85	-10.96, -7.20	-9.08

-6188.5 Avg.	**118**	**277.5**	**2.35**	**1284.53**	**-10.89**

A-allele - N+ Averages

Test #LM-MFE	Avg. Structure Energy	Avg. Barrier	Avg. Opening Energy
10	-631.2/10 -63.12	37.4/10 3.74	-118.2/10 -11.82
20	-1253.8/20 -62.69	61.9/20 3.10	-238.7/20 -11.94
62	-3844.4/62 -62.01	155.1/62 2.50	-682.99/62 -11.02
118	-7267.5/118 -61.59	277.5/118 2.35	-1284.53/118 -10.89

Ordered by Deepest Local Minima

G-Allele

Barrier	#LM	Opening	Avg. Opening
4.70	1	-21.47	-21.47
3.30	2	2*-21.47	-21.47
3.20	2	-23.35, -11.16	-17.26
3.10	3	2*-21.47, -17.70	-20.21
3.00	4 **(12)**	2*-21.47, -10.96, -7.20	-15.27
2.90	12 **(24)**	-23.35, 10*-21.47, -17.70	-21.31
2.80	7	4*-23.35, 2*-21.47, -10.83	-21.02
2.70	6	2*-22.5, 4*-21.47	-21.81
2.60	14	10*-21.47, 4*-10.20	-18.25
2.50	32 **(83)**	4*-23.55, 26*-21.47, 2*-20.21	-21.65
2.40	5	2*-21.47, 2*-19.71, -10.83	-18.64
2.30	82 **(170)**	8*-23.45, 2*-23.25, 2*-22.75, 62*-21.37, 2*-20.11, 2*-20.01, 2*-19.61, 2*-10.83	-21.29

G-Allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
12	39/12 3.25	-220.66/12 -18.39
24	73.8/24 3.08	-476.41/24 -19.85
83	226/83 2.73	-1704.3/83 -20.53
170	426.60/83 2.51	-3543.15/170 -20.84

A-Allele

Barrier	#LM	Opening	Avg. Opening
5.70	1	-7.20	-7.20
5.00	1	-17.47	-17.47
4.70	1	-17.47	-17.47
4.60	1	-5.80	-5.80
4.50	1	-8.10	-8.10
4.20	2	2*-10.40	-10.40
3.70	1	-8.30	-8.30
3.60	2 **(10)**	-10.96, -7.20	-9.08
3.40	2	2*-10.40	-10.40
3.30	5	2*-10.96, 2*-7.20, -6.90	-8.64
3.10	2	-7.20, -10.96	-9.08
3.00	4 **(23)**	-8.10, -8.03, -7.60, -7.10	-7.71
2.90	8	6*-17.47, -10.96, -7.20	-15.37
2.80	5	-9.00, 2*-7.50, -6.90, -6.83	-7.60
2.70	6	-10.96, 2*-8.30, -7.60, -7.50, -7.20	-8.31
2.60	25 **(67)**	10*-17.47, -11.90, 4*-10.96, -9.00, 2*-8.10, -7.63, 2*-7.20, 3*-5.80, -3.80	-11.95
2.50	37 **(104)**	4*-17.47, 7*-10.96, 4*-10.40, 3*-8.30, 2*-8.10, 2*-8.03, 7*-7.20, 4*-7.10, -6.60, -5.50, 2*-5.80	-9.40

A-Allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	43.8/10 4.38	-103.3/10 -10.33
23	85.3/23 3.71	-216.31/23 -9.40
67	203.7/67 3.04	-725.75/67 -10.83
104	296.2/104 2.85	-1074.01/104 -10.33

Ordered by Opening Energy

G-Allele

Opening	#LM	Barrier	Avg. Barrier
-5.40	1	1.70	1.70
-5.50	2	2.10, 1.40	1.75
-5.80	2	2*1.60	1.60
-6.20	2	2.00, 1.40	1.70
-7.20	3 **(10)**	3.00, 2.10, 1.50	1.87
-7.60	1	1.40	1.40
-8.10	2	2*1.50	1.50
-8.50	2	1.90, 1.40	1.65
-9.10	2	2*1.50	1.50
-9.20	6 **(23)**	2*1.60, 4*1.40	1.47
-9.40	2	2*1.80	1.80
-9.43	1	1.40	1.40
-9.60	4	4*1.40	1.40
-9.70	8	4*2.10, 4*1.40	1.75
-10.13	5	2.10, 1.70, 2*1.60 ,1.40	1.68
-10.20	8	4*2.60, 4*1.90	2.25
-10.80	3	1.70, 2*1.60	1.63
-10.83	8 **(62)**	2.80, 2.40, 2*2.30, 2.10, 1.70, 2*1.60	2.10
-10.96	3	3.00, 2.10, 1.50	2.20
-11.16	4	3.20, 2.10, 1.70, 1.50	2.13
-11.70	3	2.20, 1.70, 1.50	1.80
-16.30	2	2*1.70	1.70
-16.70	2	2*1.40	1.40
-17.27	2	2*1.80	1.80
-17.70	8	3.10, 2.90, 2*2.10, 2*1.60, 2*1.40	2.03
-18.07	4	4*1.60	1.60
-18.17	4	4*1.70	1.70
-18.37	6 **(100)**	4*1.90, 2*1.40	1.73

G-Allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	18.4/10 1.84	-62/10 -6.20
23	37.9/23 1.65	-176.2/23 -7.66
62	110.6/62 1.78	-571.72/62 -9.22
100	180.7/100 1.80	-1181.66/100 -11.82

A-Allele

Opening	#LM	Barrier	Avg. Barrier
-2.00	1	1.50	1.50
-2.20	1	1.40	1.40
-2.30	1	1.90	1.90
-2.60	2	2*1.40	1.40
-2.70	1	1.40	1.40
-3.00	1	1.80	1.80
-3.10	1	1.90	1.90
-3.30	5 **(13)**	2.10, 1.80, 3*1.40	1.62
-3.40	1	1.40	1.40
-3.70	1	1.40	1.40
-3.80	2	2.60, 1.90	2.25
-3.90	2	2*1.90	1.90
-4.00	4 **(23)**	2*1.50, 2*1.40	1.45
-4.10	3	2*1.50, 1.40	1.47
-4.30	2	2*1.40	1.40
-4.40	2	2*1.40	1.40
-4.50	1	1.50	1.50
-5.00	1	1.40	1.40
-5.10	3	1.70, 2*1.50	1.57
-5.40	32 **(67)**	16*1.50, 16*1.40	1.45
-5.50	4	4*1.40	1.40
-5.70	2	1.60, 1.50	1.55
-5.80	31 **(104)**	4.60, 3*2.60, 2*2.50, 2.20, 3*2.10, 2.00, 3*1.90, 4*1.80, 3*1.70, 4*1.60, 4*1.50, 2*1.40	1.97

A-Allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
13	20.8/13 1.60	-37/13 -2.85
23	37.7/23 1.52	-75.5/23 -3.28
67	101.7/67 1.52	-302.8/67 -4.52
104	171.5/104 1.65	-516/104 -4.96

4. HLA_G - (L = 386)

NM_002127.5 vs. rs1063320
miRNA: miR-148a-3p
SNP Pos. 233
Target site: 221 - 239

RNAsubopt and Barrier Setting

Offset	Barrier	C-allele	G-allele	C-allele	G-allele
4.0	1.4	9,718,256	8,406,185	11,957	10,473

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
C-allele	-23.90	-4.55	-9.28	-14.62
G-allele	-30.50	-3.24	-11.01	-19.49

Local Minima

MFE	**C-allele**		**G-allele**	
#Pairings	9		11	
Opening Energy	-12.30		-13.10	
Structure Energy	-86.70		-87.50	
MFE Barrier	4.00		4.00	
Minima				
Less Pairings	18.8%	2,248	21.0%	2,201
Equal Pairings	46.9%	5,613	55.9%	5,853
Greater Pairings	34.3%	4,095	23.1%	2,418
Identical to MFE	38.1%	4,554	45.0%	4,714
Minima (- MFE)	**11,956**		**10,472**	
Avg. Opening of All Local Minima (excluding MFE)	**-10.24**		**-11.04**	

Less/Equal Pairings than MFE Target Site

#Pairings	#Minima C-allele		Target Site Approx. Avg.	#Minima G-allele		Target Site Approx. Avg.
6	3.00%	359	-6.88	0.72%	75	-6.30
7	7.89%	943	-5.15	7.26%	760	-5.79
8	7.91%	946	-7.19	7.65%	801	-8.08
9	**46.95%**	**5,613**	**-10.96**	1.48%	155	-7.49
10	**2.18%**	**261**	**-4.80**	3.92%	410	-7.29
11	**0.92%**	**110**	**-6.56**	**55.89%**	**5,853**	**-11.78**
Total Minima	**68.85%**	**8,232**	**-9.43**	**76.91%**	**8,054**	**-10.48**
Avg. Barrier	**1.71**			**1.71**		

Local Minima with Less Pairings

	C-allele	G-allele
#LM Less Pairings:	18.8% 2,248	21.0% 2,201
Avg. Target Site Energy	-6.28	-7.04
Less Pairings Avg. Barrier Height	1.75	1.77

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	C-allele	G-allele
Less	18.8% 6,448	48.8% 5,112
Equal	35.7% 4,272	41.7% 4,372
Less or Equal	**89.7% 10,720**	**90.6% 9,484**

N+ Test

C-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-86.00	33 **(112)**	4*2.30, 4*3.30, 2*3.00, 8*2.80, 2*2.70, 8*2.50, 2.20, 4*1.40	2.55	16*-12.30, 4*-11.60, 8*-6.90, 2*-4.80, 3*-4.10	-9.71
-86.10	24 **(79)**	3.40, 2.90, 2*2.70, 4*2.50, 8*2.30, 8*1.40	2.14	12*-12.30, 8*-6.90, 2*-4.80, 2*-4.10	-9.19
-86.20	17	3*3.50, 3.30, 2*3.00, 2*2.90, 8*2.30, 2.70	2.75	12*-12.30, 3*-4.80, 2*-4.10	-10.01
-86.30	13	2*3.60, 2*3.30, 2.90, 8*2.80	3.00	4*-12.30, 4*-6.90, 2*-4.80, 3*-4.10	-7.59
-86.40	11 **(25)**	4*3.70, 3.30, 2*3.00, 4*2.50	3.10	8*-12.30, -4.80, 2*-4.10	-10.13
-86.50	10 **(14)**	2*3.80, 4*2.30, 4*1.40	2.24	8*-12.30, -4.80, -4.10	-10.73
-86.60	1	3.90	3.90	-4.10	-4.10
-86.70	3	4.00, 2*3.00	3.33	3*-12.30	-12.30
MFE: -86.70	1	4.00	4.00	-12.30	-12.30
-9740.2 Avg. -86.20	**113**	**295.5**	**2.62**	**1081.8**	**-9.68**

C-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
112	-9653.8/112 -86.19	291.5/112 2.60	-1069.5/112 -9.55
79	-6815.8/79 -86.28	207.5/79 2.63	-749.2/79 -9.48
25	-2162.1/25 -86.48	70.4/25 2.82	-259.7/25 -10.39
14	-1211.7/14 -86.55	36.3/14 2.59	-148.3/14 -10.59

N+ Test

G-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-86.80	30 **(118)**	3*3.30, 2.90, 2*2.70, 4*2.30, 8*2.50, 8*2.80, 4*1.40	2.51	16*-13.10, -11.60, -11.20, 8*-7.70, 2*-5.50, 2*-4.80	-10.49
-86.90	27	3*3.40, 3.30, 2*2.90, 2.70, 4*2.50, 8*2.30, 8*1.40	2.28	12*-13.10, -11.60, -11.20, 8*-7.70, 3*-5.50, 2*-4.80	-9.54
-87.00	27 **(61)**	6*3.50, 2*3.30, 2*3.00, 2.90, 8*2.30, 8*1.40	2.45	20*-13.10, -11.60, -11.20, 2*-5.50, 3*-4.80	-11.49
-87.10	12	3*3.60, 3.30, 8*2.80	3.04	4*-13.10, -11.60, 4*-7.70, -5.50, 2*-4.80	-9.16
-87.20	10 **(22)**	4*3.70, 2*3.00, 4*2.50	3.08	8*-13.10, -5.50, -4.80	-11.51
-87.30	9 **(12)**	3.80, 4*2.30, 4*1.40	2.07	8*-13.10, -4.80	-12.18
-87.50	3	4.00, 2*3.00	3.50	3*-13.10	-13.10
MFE: -87.50	1	4.00	4.00	-13.10	-13.10
-10352.2 **Avg.** **-86.99**	**119**	**303.00**	**2.55**	**1279.5**	**-10.75**

G-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
118	-10264.7/118 -86.99	299/118 2.53	-1266.4/118 -10.73
61	-5314.4/61 -87.12	162/61 2.66	-684.1/61 -11.21
22	-1920.2/22 -87.28	59.4/22 2.70	-264.22/22 -12.00
12	-1848.2/12 -87.35	29.13/12 2.42	148.9/12 12.41

Ordered by Deepest Local Minima

C-allele

Barrier	#LM	Opening	Avg. Opening
4.00	1	-12.30	-12.30
3.90	1	-4.10	-4.10
3.80	2	-4.10, -4.80	-4.45
3.70	4	-4.10, -4.80, 2*-12.30	-8.38
3.60	2 **(10)**	-4.10, -4.80	-4.45
3.50	3	-4.80, 2*-12.30	-9.80
3.40	1	-4.10	-4.10
3.30	8 **(22)**	3*-4.10, 3*-4.80, 2*-11.60	-6.24
3.20	16	6*-4.10, -4.80, 9*-12.30	-8.76
3.10	11	4*-4.10, 4*-4.0, -7.20, 2*-18.00	-6.87
3.00	37 **(86)**	-4.10, 4*-4.80, -7.20, 4*-9.80, 6*-11.60, 19*-12.30, 2*-18.00	-11.05
2.90	30 **(116)**	6*-4.10, 5*-4.80, 4*-6.30, 3*-7.20, 12*-12.30	-8.10

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	37.5/10 3.75	-67.7/10 -6.77
22	77.8/22 3.54	-151.1/22 -6.87
86	274.1/86 3.19	-779/86 -9.06
116	361.10/116 3.11	-1022/116 -8.81

G-allele

Barrier	#LM	Opening	Avg. Opening
4.00	1	-13.10	-13.10
3.80	1	-4.80	-4.80
3.70	4	-4.80, -5.50, 2*-13.10	-9.13
3.60	3 **(9)**	-4.80, -5.50, -11.60	-7.30
3.50	6	-4.80, -5.50, -11.20, -11.60, 2*-13.10	-9.88
3.40	3	-5.50, -11.20, -11.60	-9.43
3.30	7 **(25)**	3*-4.80, 2*-5.50, -11.20, -11.60	-6.88
3.20	15	-4.80, -5.50, -11.20, 2*-12.30, 9*-13.10, -15.71	-11.98
3.10	8	6*-4.80, -5.50, -15.71	-6.25
3.00	33 **(81)**	4*-4.80, 4*-5.50, -6.80, 2*-12.30, 19*-13.10, 3*-15.71	-11.17
2.90	40 **(121)**	5*-4.80, 7*-5.50, -6.80, 4*-7.00, -7.50, 4*-12.30, 12*-13.10, 4*-15.71, 2*-15.22	-10.11

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
9	33.4/9 3.71	-76.3/9 -8.48
25	87.7/25 3.51	-212.1/25 -8.48
81	259.5/81 3.20	-810.45/81 -10.01
121	375.50/121 3.10	-1214.93/121 -10.04

Ordered by Opening Energy

C-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-3.10	4	2*1.60, 2*1.50	1.55
-3.30	29 **(33)**	2.10, 2.00, 3*1.90, 4*1.80, 4*1.70, 4*1.60, 4*1.50, 8*1.40	1.63
-4.00	21 **(65)**	2.00, 1.90, 3*1.80, 4*1.70, 4*1.60, 4*1.50, 4*1.40	1.62
-4.10	495 **(560)**	81*1.40, 112*1.50, 39*1.60, 61*1.70, 29*1.80, 30*1.90, 17*2.00, 14*2.10, 29*2.20, 11*2.30, 10*2.40, 9*2.50, 8*2.60, 19*2.70, 2.80, 6*2.90, 3.00, 4*3.10, 6*3.20, 3*3.30, 3.40, 3.60, 3.70, 3.80, 3.90	1.84

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	17.8/10 1.78	-32.2/10 -3.22
33	53.6/33 1.62	-108.1/33 -3.28
65	103.1/65 1.59	-237.2/65 -3.65
560	1014/560 1.81	-2274.4/112 -4.06

G-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-3.80	4	2*1.50, 2*1.40	1.45
-4.00	21	2.00, 1.90, 3*1.80, 4*1.70, 4*1.60, 4*1.50, 4*1.40	1.62
-4.70	17	1.90, 1.80, 3*1.70, 4*1.60, 4*1.50, 4*1.40	1.58
-4.80	418	3.80, 3.70, 3.60, 3.50, 3*3.30, 3.20, 6*3.10, 4*3.00, 5*2.90, 2*2.80, 10*2.70, 10*2.60, 10*2.50, 7*2.40, 10*2.30, 26*2.20, 14*2.10, 14*2.00, 24*1.90, 23*1.80, 58*1.70, 32*1.60, 88*1.50, 67*1.40	1.85

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	16.8/10 1.68	-39.2/10 -3.92

4. HLA_G (L = 386)

25	39.9/25 1.60	-99.2/25 -3.97
60	125/60 2.08	-265.5/60 -4.43
460	840.10/460 1.83	-2185.50/460 -4.75

5. PARP1 (L = 769)

NM_001618.3 vs. rs8679
miRNA: miR-145-5p
SNP Pos. 607
Target site: 592-614

RNAsubopt and Barrier Setting

Offset	Barrier	U-allele	C-allele	U-allele	C-allele
6.0	1.4	10,610,542	1,582,333	14,281	1,709

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
U-allele	-19.70	0.93	-12.19	-7.51
C-allele	-19.20	-0.65	-20.90	1.70

Local Minima

MFE	**U-allele**		**C-allele**	
#Pairings	11		22	
Opening Energy	-18.10		-31.12	
Structure Energy	-186.90		-188.70	
MFE Barrier	3.00		3.00	
Minima				
Less Pairings	79.1%	11,295	12.3%	210
Equal Pairings	7.9%	1,226	87.7%	1,499
Greater Pairings	12.3%	1,760	0.0%	0
Identical to MFE	8.6%	1,226	42.2%	807
Minima (- MFE)	**14,281**		**1,709**	
Avg. Opening of All Local Minima (excluding MFE)	**-12.83**		**-30.81**	

Less/Equal Pairings than MFE Target Site

#Pairings	#Minima U-allele		Target Site Approx. Avg.	#Minima C-allele		Target Site Approx. Avg.
6	52.06%	7,434	-8.75	-		-
7	21.70%	3,099	-10.19	-		-
10	5.34%	762	-17.90	-		-
11	**8.58%**	**1,226**	**-18.10**	-		-
12	**0.01%**	**1**	**-19.20**	-		-
17	**0.01%**	**2**	**-22.40**	-		-
20	**1.56%**	**223**	**-27.98**	12.17%	208	-30.46
21	**0.01%**	**2**	**-25.50**	0.12%	2	-27.40
22	**10.73%**	**1,532**	**-28.97**	**87.71%**	**1,499**	**-30.87**
Total Minima	100%	14,281	-12.83	100%	1,709	-30.81

Local Minima with Less Pairings

	U-allele	C-allele
#LM Less Pairings:	79.1% 11,295	12.29% 210
Avg. Target Site Energy	-9.76	-30.43
Less Pairings Avg. Barrier Height	1.42	1.51

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	U-allele	C-allele
Less	79.1% 11,295	12.3% 902
Equal	8.6% 1,226	42.2% 807
Less or Equal	**87.7% 12,521**	**100% 1,709**

N+ Test

U-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-186.30	48 **(103)**	24*2.40, 10*2.00, 10*1.70, 4*1.50	2.10	12*-28.72, 10*-18.10, 2*-17.90, 8*-10.20, 16*-8.80	-16.33
-186.40	14 **(55)**	12*2.50, 2*2.00	2.43	8*-18.10, 6*-17.90	-18.01
-186.50	18	10*2.60, 4*2.40, 4*2.00	2.42	4*-29.22, 2*-18.10, 4*-10.20, 8*-8.80	-14.68
-186.60	12 **(23)**	6*2.70, 2*2.60, 2*2.00, 2*1.20	2.32	4*-29.22, 6*-18.10, 2*-17.90	-12.77
-186.70	6 **(11)**	2*2.80, 4*2.40	2.53	4*29.22, 2*-17.90	-25.45
-186.80	4	2.90, 2.70, 2*2.60	2.70	4*-29.22	-29.22
-186.90	1	3.00	3.00	-18.10	-18.10
MFE: -186.90	1	3.00	3.00	-18.10	-18.10
-19389.4 Avg. -186.44	**104**	**238**	**2.29**	**1867.36**	**-17.96**

U-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
103	-19202.5/103 -186.43	235/103 2.28	-1849.26/103 -17.95
55	-1065.42/55 -186.55	134.4/55 2.44	-1065.42/55 -19.37
23	-4293.5/23 -186.67	56.8/23 2.47	-548.94/23 -23.87
11	-2054.3/11 -186.75	29/11 2.64	-287.66/11 -26.15

N+ Test

C-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-187.90	40 **(123)**	20*2.20, 12*2.10, 4*1.70, 4*1.50	2.05	4*-30.52, 12*-30.62, 24*-31.12	-30.91
-188.00	32 **(83)**	20*2.30, 4*2.10, 4*1.70, 4*1.50	2.10	4*-30.52, 16*-31.12, 12*-30.62	-30.86
-188.10	24	12*2.40, 4*2.10, 4*1.70, 4*1.50	2.08	12*-30.62, 4*-30.52, 8*-31.12	-30.77
-188.20	12 **(27)**	4*2.50, 4*1.70, 4*1.50	1.90	12*-30.62	-30.62
-188.40	4	4*2.40	2.40	4*-31.12	-31.12
-188.50	4 **(11)**	2.80, 2.70, 2*2.60	2.68	4*-31.12	-31.12
-188.60	4	4*2.40	2.40	4*-31.12	-31.12
-188.70	3	2.70, 2*2.60	2.63	3*-31.12	-31.12
MFE: -188.70	1	3.00	3.00	-31.12	-31.12
-23321.6 Avg. -188.08	**124**	**262.8**	**2.12**	**3827.68**	**-30.87**

C-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
123	-23132.9/123 -188.07	259.8/123 2.11	-3796.56/123 -30.87
83	-15616.9/83 -188.16	117.8/83 2.14	-2560.16/83 -30.85
27	-5086.5/27 -188.39	60.6/27 2.24	-834.24/27 -30.90
11	-2074.5/11 -188.6	28.2/11 2.56	342.32/11 -31.12

Ordered by Deepest Local Minima

U-allele

Barrier	#LM	Opening	Avg. Opening
3.00	1	-18.10	-18.10
2.90	1	-29.22	-29.22
2.80	2	2*-17.90	-17.90
2.70	7 **(11)**	4*-18.10, 3*-29.22	-22.87
2.60	14 **(25)**	4*-8.80, 4*-10.20, 2*-18.10, 4*-29.22	-16.36
2.50	12	4*-17.90, 8*-18.10	-18.03
2.40	32 **(69)**	6*-8.80, 6*-10.20, 2*-17.90, 6*-18.10, 4*-28.72, 8*-29.22	-18.97
2.30	54 **(123)**	4*-8.50, 12*8.80, 12*-10.20, 8*-17.90, 6*-18.10, 4*-28.72, 8*-29.22	-15.97

U-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
11	30.4/11 2.76	-243.18/11 -22.11
25	66.8/25 2.67	-472.26/25 -18.89
69	173.6/69 2.52	-1295.7/69 -18.78
123	297.8/123 2.42	-2158.14/123 -17.55

C-allele

Barrier	#LM	Opening	Avg. Opening
2.80	1	-31.12	-31.12
2.70	2	2*-31.12	-31.12
2.60	4	4*-31.12	-31.12
2.50	4 **(11)**	4*-30.62	-30.62
2.40	20 **(31)**	16*-31.12, 4*30.62	-31.02
2.30	20	16*-31.12, 4*-30.62	-31.02
2.20	20 **(71)**	16*-31.12, 4*-30.62	-31.02
2.10	64 **(135)**	16*-30.52, 48*-31.12	-30.97

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
11	28.6/11 2.60	-340.32/11 -30.94
31	76.6/31 2.47	-960.72/31 -30.99
71	166.6/71 2.35	-2201.52/71 -31.01
135	301/135	-4183.60/116

	2.23	-30.99

Ordered by Opening Energy

U-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-8.50	1,300	4*2.30, 6*2.10, 16*2.00, 13*1.90, 48*1.80, 109*1.70, 64*1.60, 161*1.50, 183*1.40, 213*1.30, 483*1.20	1.39
-8.80	6,134	4*2.60, 6*2.40, 12*2.30, 10*2.20, 32*2.10, 172*2.00, 166*1.90, 230*1.80, 568*1.70, 378*1.60, 748*1.50, 760*1.40, 934*1.30, 2114*1.20	1.42

U-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	21.8/10 2.18	-85/10 -8.50
20	41.8/20 2.09	-170/20 -8.50
60	116.3/60 1.94	-510/60 -8.50
100	187/100 1.87	-850/100 -8.50

C-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-27.40	2	2*1.30	1.30
-27.50	4	2*1.40, 2*1.30	1.35
-30.02	24	16*1.50, 8*1.30	1.43
-30.12	40	16*1.50, 8*1.40, 16*1.30	1.40
-30.52	184	40*2.10, 24*1.60, 40*1.50, 44*1.40, 36*1.30	1.58

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	14/10 1.40	-284.88/10 -28.49
20	29/20 1.45	-585.08/20 -29.54
70	98.4/70 1.41	-2090.08/70 -29.86
254	378.8/254 1.49	-7705.76/254 -30.34

6. WFS1 (L = 797)

NM_001145853.1 vs. rs1046322
miRNA: miR-668-3p
SNP Pos. 253
Target site: 234 - 258

RNAsubopt and Barrier Setting

Offset	Barrier	G-allele	A-allele	G-allele	A-allele
2.7	0.8	11,404,145	11,362,245	79,577	79,273

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
G-allele	-26.60	-4.421	-10.641	-15.959
A-allele	-20.60	-1.723	-14.196	-6.404

Local Minima

MFE	**G -allele**	**A-allele**
#Pairings	12	12
Opening Energy	-15.72	-15.72
Structure Energy	-290.70	-290.70
MFE Barrier	2.70	2.70
Minima		
Less Pairings	0.34% 272	0%
Equal Pairings	92.9% 73,943	93.3% 73,943
Greater Pairings	6.74% 5,360	6.7% 5,328
Target Site Identical to MFE	92.9% 73,943	93.3% 73,943
Minima (- MFE)	**79,575**	**79,271**
Avg. Opening of All Local Minima (excluding MFE)	**-15.64**	**-15.67**

Less/Equal Pairings than MFE

#Pairings	#Minima G-allele	Target Site Approx. Avg.	#Minima A-allele	Target Site Approx. Avg.
10	0.34% 272	-7.56	0	-
Total:	**0.34% 272**	**-7.56**	**0**	**-**
Avg. Barrier:	**1.01**		**-**	

Local Minima with Less or Equal Approx. Target Site Energy than the MFE

	G-allele	A-allele
Less	7% 5,600	6.7% 5,328
Equal	92.9% 73,943	93.3% 73,943
Less or Equal	99.96% 79,543	100% 79,271

N+ Test

G-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-290.50	96	16*2.50, 16*2.30, 32*1.90, 32*1.30	1.87	96*-15.72	-15.72
-290.60	64	32*2.60, 16*1.90, 16*1.30	2.10	64*-15.72	-15.72
MFE: -290.70	16	16*2.70	2.70	16*-15.72	-15.72

G-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
10	-290.70	2.70	-15.72
20	-290.68	2.50	-15.72
60	-290.63	2.44	-15.72
100	-290.59	2.15	-15.72

N+ Test

A-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-290.50	96	16*2.50, 16*2.30, 32*1.90, 32*1.30	1.87	96*-15.72	-15.72
-290.60	64	32*2.60, 16*1.90, 16*1.30	2.10	64*-15.72	-15.72
MFE: -290.70	16	16*2.70	2.70	16*-15.72	-15.72

G-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
10	-290.70	2.70	-15.72
20	-290.68	2.50	-15.72
60	-290.63	2.44	-15.72
100	-290.59	2.15	-15.72

Ordered by Deepest Local Minima

G-Allele

Barrier	#LM	Opening	Avg. Opening
2.70	15	-15.72	-15.72
2.60	32	-15.72	-15.72
2.50	16	-15.72	-15.72
2.40	16	-15.72	-15.72
2.30	80	-15.72	-15.72

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	2.70	-15.72
20	2.68	-15.72
60	2.65	-15.72
100	2.50	-15.72

A-Allele

Barrier	#LM	Opening	Avg. Opening
2.70	15	-15.72	-15.72
2.60	32	-15.72	-15.72
2.50	16	-15.72	-15.72
2.40	16	-15.72	-15.72
2.30	80	-15.72	-15.72

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	2.70	-15.72
20	2.68	-15.72
60	2.65	-15.72
100	2.50	-15.72

Ordered by Opening Energy

G-allele

Opening	#LM	Barrier	Avg. Barrier	Structure Energy	Avg. Structure
-7.30	32	32*0.90	0.90	-288.90	-288.90
-7.60	240	32*1.20, 64*1.10, 64*1.00, 80*0.90	1.02	32*-289.20 64*-289.10 64*-289.00 80*-288.90	-289.02

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	0.90	-7.30
20	0.90	-7.30
60	0.90	32*-7.30, 28*-7.60 Avg. -7.44
100	0.90	32*-7.30, 68*-7.60 Avg. -7.50

A-allele

Opening	#LM	Barrier	Avg. Barrier	Structure Energy	Avg. Structure
-14.92	5,328	96*0.80 1,544*0.90 1,400*1.00 1040*1.10 680*1.20 344*1.30 80*1.50 64*1.60 64*1.70 16*1.90	1.06	1544*-288.90 1,400*-289.00 1088*-289.10 552*-289.20 376*-289.40 96*-289.50 96*-289.60 96*-289.70 16*-289.80 64*-289.90	-289.09

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	0.90	-14.92
20	0.90	-14.92
60	0.90	-14.92
100	0.90	-14.92

7. IL23R (L = 851)

NM_144701.2 vs. rs10889677
miRNA: let-7e
SNP Pos. 309

	C-allele	A-allele
Target Site:	291 - 310	291 - 308

RNAsubopt and Barrier Setting

Offset	Barrier	C-allele	A-allele	C-allele	A-allele
2.0	1.2	396,014	756,122	964	1,080

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
C-allele	-24.70	0.00	-25.82	1.12
A-allele	-20.70	-0.05	-21.40	0.70

Local Minima

MFE	C-allele		A-allele	
#Pairings	17		15	
Opening Energy	-25.90		-21.70	
Structure Energy	-228.40		-224.60	
MFE Barrier	2.00		2.00	
Minima				
Less Pairings	0.0%	0	35.6%	384
Equal Pairings	100%	963	64.4%	695
Greater Pairings	0.0%	0	0.0%	0
Identical to MFE	100%	963	64.4%	695
Minima (- MFE)	**963**		**1,079**	
Avg. Opening of All Local Minima (excluding MFE)	**-25.90**		**-18.46**	

#Pairings	#Minima C-allele		Target Site Approx. Avg.	#Minima A-allele		Target Site Approx. Avg.
12	-		-	29.7%	320	-12.50
13	-		-	5.9%	64	-13.10
14	-		-	-		-
15	-		-	**64.4%**	**695**	**-21.70**
16	-		-	-		-
17	**100%**	**963**	**-25.90**	-		-
Total Minima	**100%**	**963**	**-25.90**	**100%**	**1,079**	**-18.46**
Avg. Barrier	**1.37**			**1.44**		

Local Minima with Less Pairings

	C-allele	A-allele
#LM Less Pairings:	0.0%	35.6% 384
Avg. Target Site Energy	-	-12.60
Less Pairings Avg. Barrier Height	-	1.40

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	C-allele		A-allele	
Less	0.0%	0	35.6%	384
Equal	100%	963	64.4%	695
Less or Equal	100%	963	100%	1,079

N+ Test

C-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-228.10	40 **(135)**	40*1.70	1.70	40*-25.90	-25.90
-228.20	32	32*1.70	1.70	32*-25.90	-25.90
-228.30	32 **(63)**	32*1.90	1.90	32*-25.90	-25.90
-228.40	31 **(31)**	31*2.00	2.00	31*25.90	-25.90
MFE -228.40	1	2.00	2.00	-25.90	-25.90
31040.8 Avg. -228.24	**136**	**247.2**	**1.80**	**3522.4**	**-25.90**

C-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
135	-30812.4/135 -228.24	245.2/135 1.86	-3496.5/135 -25.90
63	-14386/63 -228.35	122.8/63 1.95	-1631.7/63 -25.90
31	-7080.4/31 -228.4	62/31 2.00	-802.9/31 -25.90
10	-2284/10 -228.4	20/10 2.00	-259.00 -25.90

N+ Test

A-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-224.30	40 **(135)**	40*1.70	1.70	40*-21.70	-21.70
-224.40	32	32*1.70	1.70	32*-21.70	-21.70
-224.50	32 **(63)**	32*1.90	1.90	32*-21.70	-21.70
-224.60	31 **(31)**	31*2.00	2.00	31*-21.70	-21.70
MFE: -224.60	1	2.00	2.00	-21.70	-21.70
-30524 Avg. -224.44	**136**	**247.2**	**1.82**	**2951.2**	**-21.70**

A-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
135	-30299.4/135 -224.44	245.2/135 1.82	-2929.5/135 -21.70
63	-14146.6/63 -224.55	122.8/63 1.95	-1367.1/63 -21.70
31	-6962.6/31 -224.60	62/31 2.00	672.7/31 -21.70
10	-2246/10 -224.60	20/10 2.00	-217/10 -21.70

Ordered by Deepest Local Minima

C-allele

Barrier	#LM	Opening	Avg. Opening
2.00	31	31*-25.90	-25.90
1.90	32	32*-25.90	-25.90
1.80	72	72*-25.90	-25.90

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	20/10 2.00	-259/10 -25.90
20	40/20 2.00	-518/20 -25.90
60	117.1/60 1.95	-1554/69 -25.90
100	189.4/100 1.89	-2590/100 -25.90

A-allele

Barrier	#LM	Opening	Avg. Opening
2.00	31	31*-21.70	-21.70
1.90	32	32*-21.70	-21.70
1.70	72	72*-21.70	-21.70

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	20/10 2.00	-217/10 -21.70
20	40/20 2.00	-434/20 -30.99
60	117.1/60 1.95	-1302/60 -21.70
100	185.7/100 1.86	-2170/100 -21.70

Ordered by Opening Energy

C-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-25.90	963	31*2.00, 32*1.90, 72*1.70, 56*1.50, 232*1.40, 200*1.30, 340*1.20	1.37

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	20/10 2.00	-259/10 -25.90
20	40/20 2.00	-518/20 -25.90
60	117.1/60 1.95	-1554/60 -25.90
100	185.7/100 1.86	-2590/100 -25.90

A-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-12.50	320	128*1.30, 64*1.40, 64*1.50, 64*1.60	1.42

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	16/10 1.60	-125/10 -12.50
20	32/20 1.60	-250/20 -12.50

60	72/60 1.60	-750/60 -12.50
100	156.4/100 1.56	-1250/100 -12.50

8. RYR3 (L = 880)

NM_001036.3 vs. rs1044129
miRNA: miR-367
SNP Pos: 839

	A-allele	G-allele
Target Site:	835 - 857	830 - 857

RNAsubopt and Barrier Setting

Offset	Barrier	A-allele	G-allele	A-allele	G-allele
2.3	1.2	13,569,252	5,006,010	19,936	2,628

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
A-allele	-14.80	-1.22	-11.42	-3.38
G-allele	-15.30	-0.19	-19.65	4.35

Local Minima

MFE	A-allele		G-allele	
#Pairings	7		14	
Opening Energy	-10.85		-11.60	
Structure Energy	-186.60		-188.40	
MFE Barrier	2.30		2.30	
Minima				
Less Pairings	1.6%	324	0.0%	0
Equal Pairings	56.3%	11,227	96.6%	2,537
Greater Pairings	42.1%	8,384	3.4%	90
Identical to MFE	56.2%	11,203	90.6%	2,379
Minima (- MFE)	**19,935**		**2,627**	
Avg. Opening of All Local Minima (excluding MFE)	**-10.93**		**-11.75**	

Local Minima

#Pairings	#Minima A-allele		Target Site Approx. Avg.	#Minima G-allele	Target Site Approx. Avg.
6	1.6%	324	-9.19	-	-
7	**56.3%**	**11,227**	**-10.85**	-	-
8	**36.7%**	**7,322**	**-11.09**	-	-
9	**5.3%**	**1,062**	**-11.28**	-	-
10	-		-	-	-
11	-		-	-	-
12	-		-	-	-
13	-		-	-	-
14	-		-	**2,537**	**-11.57**
Total Minima	**19,935**		**-10.93**	**2,537**	**-11.57**
Avg. Barrier	**1.42**			**1.44**	

Local Minima with Less Pairings

	A-allele	G-allele
#LM Less Pairings:	1.6% 324	0
Avg. Target Site Energy	-9.19	-
Less Pairings Avg. Barrier Height	1.37	-

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	A-allele	G-allele
Less	19.1% 3,816	5.3% 140
Equal	56.2% 11,203	90.6% 2,379
Less or Equal	**73.3% 15,019**	**95.9% 2,519**

N+ Test

A-allele - 100 Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-186.40	90 **(185)**	24*2.10, 12*2.00, 12*1.80, 12*1.50, 30*1.40	1.73	30*-11.58, 60*-10.85	-11.09
-186.50	72 **(95)**	24*2.20, 8*2.00, 8*1.80, 8*1.50, 24*1.40	1.79	12*-11.58, 60*-10.85	-10.97
-186.60	23 **(23)**	15*2.30, 8*1.40	1.99	23*-10.85	-10.85
MFE: -186.60	1	2.30	2.30	-10.85	-10.85
-34682.4 Avg. -186.46	**186**	**332.8**	**1.79**	**2048.76**	**-11.01**

A-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
10	-1866/10 -186.60	23/10 2.30	-108.5/10 -10.85
23	-4291.8/23 -186.60	45.7/23 1.99	-249.55/23 -10.85
95	-17719.8/95 -186.52	174.5/95 1.84	-1039.5/95 -10.94
185	-34495.8/185 -186.46	330.5/185 1.79	-2037.9/185 -11.02

N+ Test

G-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-188.00	62 **(139)**	34*1.90, 4*1.80, 4*1.60, 20*1.40	1.71	62*-11.60	-11.60
-188.10	48 **(77)**	24*2.00,4*1.80, 4*1.60, 16* 1.40	1.75	48*-11.60	-11.60
-188.20	18 **(29)**	8*2.10, 4*1.60, 6*1.40	1.76	18*-11.60	-11.60
-188.30	6	4*2.20, 2*1.40	1.93	6*-11.60	-11.60
-188.40	5	3*2.30, 2*1.40	1.94	5*-11.60	-11.60
MFE: -188.40	1	2.30	2.30	-11.60	-11.60
-26332.6 Avg. -188.09	**140**	**245.4**	**1.75**	**-1624**	**-11.60**

G-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
11	-2071.8/11 -188.35	21.3/11 1.94	127.6/11 -11.60
29	-5459.4/29 -188.26	52.9/29 1.82	-336.4/29 -11.60
77	-14488.2/77 -188.16	136.9/77 1.78	-893.2/77 -11.60
139	-26144.2/139 -188.09	243.1/139 1.75	-1612.4/139 -11.60

Ordered by Deepest Local Minima

A-allele

Barrier	#LM	Opening	Avg. Opening
2.30	15	15*-10.85	-10.85
2.20	24	24*-10.85	-10.85
2.10	24	24*-10.85	-10.85
2.00	136	104*-10.85, 32*-11.58	-11.02

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	23/10 2.30	-108.5/10 -10.85
20	45.5/20 2.28	-217/20 -10.85
63	137.7/63 2.19	-683.55/63 -10.85

199	409.7/100 2.06	-2182.51/100 -10.97

G-allele

Barrier	#LM	Opening	Avg. Opening
2.30	3	3*-11.60	-11.60
2.20	4	4*-11.60	-11.60
2.10	8	8*-11.60	-11.60
2.00	24	24*-11.60	-11.60
1.90	34	34*-11.60	-11.60
1.80	62	62*-11.60	-11.60

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	22/10 2.20	-116/10 -11.60
20	42.5/20 2.13	-232/20 -11.60
73	145.1/60 1.99	-846.8/73 -11.60
135	256.7/100 1.90	-1566/135 -11.60

Ordered by Opening Energy

A-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-8.39	24	24*1.30	1.30
-9.19	324	16*1.60, 40*1.50, 88*1.40, 180*1.30	1.37

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	13/10 1.30	-83.9/10 -8.39
20	26/20 1.30	-167.8/20 -8.39
60	86.8/60 1.45	-532.2/60 -8.87
100	144.8/100 1.45	-899.8/100 -9.00

G-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-11.00	140	4*1.70, 4*1.60, 12*1.50, 58*1.40, 62*1.30	1.38

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	16.2/10 1.62	-110/10 -11.00
20	31.2/20 1.56	-220/20 -11.00
60	87.2/60 1.45	-660/60 -11.00
100	141/100 1.41	-1100/100 -11.00

9. AGTR1 (L = 888)

NM_032049.3 vs. rs5186
miRNA: miR-155-5p
SNP Pos: 86

	A-allele	C-allele
Target Site:	57 - 90	57 - 90

RNAsubopt and Barrier Setting

Offset	Barrier	A-allele	C-allele	A-allele	C-allele
2.3	0.8	546,284	2,147,815	2,309	14,944

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
A-allele	-20.90	-0.577	-28.561	7.661
C-allele	-16.60	-2.268	-27.879	11.279

Local Minima

MFE	A -allele	C-allele
#Pairings	22	24
Opening Energy	-18.60	-21.30
Structure Energy	-188.00	-188.80
MFE Barrier	2.30	2.30
Minima		
Less Pairings	8.1% 186	0%
Equal Pairings	91.9% 2,122	100% 14,943
Greater Pairings	0%	0%
Target Site Identical to MFE	45.8% 1,058	25% 3,735
Minima (- MFE)	**2,308**	**14,943**
Avg. Opening of All Local Minima (excluding MFE)	**-18.54**	**-21.30**

Less/Equal Pairings than MFE

#Pairings	#Minima A-allele	Target Site Approx. Avg.	#Minima C-allele	Target Site Approx. Avg.
18	8.1% 186	-17.90	0	-
22	91.9% 2,122	-18.60	0	-
24	0	-	100% 14,943	-21.30
Total:	**2,308**	**-18.54**	**14,943**	**-21.30**
Avg. Barrier:	**1.01**		**1.00**	

Local Minima with Less Pairings

	A-allele	C-allele
#LM Less Pairings:	8.1% 186	0%
Avg. Target Site Energy	-17.90	-
Less Pairings Avg. Barrier Height	0.97	-

Local Minima with Less or Equal Approx. Target Site Energy than the MFE

	A-allele	C-allele
Less	8.3% 191	0%
Equal	91.7% 2,117	100% 14,943
Less or Equal	100% 2,308	100% 14,943

N+ Test

A-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-187.30	44 **(122)**	30*1.60, 10*1.40, 4*0.90	1.49	44*-18.60	-18.60
-187.31	2	2*1.61	1.61	2*-18.60	-18.60
-187.40	22 **(76)**	10*1.70, 10*1.40, 2*0.90	1.49	22*-18.60	-18.60
-187.41	6	3*1.71, 2*1.60	1.39	6*-18.60	-18.60
-187.50	14	10*1.80, 2*1.60, 2*1.40	1.71	14*-18.60	-18.60
-187.51	2	2*1.81	1.81	2*-18.60	-18.60
-187.60	10	4*1.90, 6*1.40	1.60	110*-18.60	-18.60
-187.61	2	2*1.91	1.91	2*-18.60	-18.60
-187.70	8 **(20)**	4*2.00, 2*1.60, 2*1.40	1.75	8*-18.60	-18.60
-187.71	2	2*2.01	2.01	2*-18.60	-18.60
-187.80	4 **(10)**	2*2.10, 2*1.40	1.75	4*-18.60	-18.60
-187.90	4	2*2.20, 2*1.40	1.80	4*-18.60	-18.60
-188.00 MFE	2	2*2.30	2.30	2*-18.60	-18.60
Total: -22870.14 Avg. -187.46	**122**	**194.21**	**1.59**	**-2269.2**	**-18.60**

A-allele - N+ Averages

Test #LM-MFE	Avg. Structure Energy	Avg. Barrier	Avg. Opening Energy
10	-1878.8/10 -187.88	18.8/10 1.88	-18.60
20	-3755.82/20 -187.79	36.82/20 1.84	-18.60
76	-14254.32/76 187.56	125.39/76 1.65	-18.60
122	-22870.14/122 -187.46	194.21/122 1.59	-18.60

N+ Test

C-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-188.50	48 **(120)**	16*2.00, 8*1.80, 12*1.60, 12*1.40	1.72	48*-21.30	-21.30
-188.51	12	8*2.01, 4*1.80	1.94	12*-21.30	-21.30
-188.60	24 **(60)**	8*2.10, 4*1.80, 12*1.40	1.70	24*-21.30	-21.30
-188.70	24 **(36)**	8*2.20, 4*1.80, 12*1.40	1.73	36*-21.30	-21.30
-188.80 MFE	12 **(12)**	8*2.30, 4*1.80	2.13	12*-21.30	-21.30
Total: -22630.92 Avg. -188.59	**120**	**191.4/120**	**1.60**	**-2556/120**	**-21.30**

C-allele - N+ Averages

Test #LM-MFE	Avg. Structure Energy	Avg. Barrier	Avg. Opening Energy
12	-2265.6/12 -188.80	26.6/12 2.13	-21.30
36	-6794.4/36 -188.73	68.2/36 1.89	-21.30
60	-11320.8/60 -188.68	109/60 1.82	-21.30
120	-22630.92/120 -188.59	191.4/120 1.60	-21.30

Ordered by Deepest Local Minima

A-Allele

Barrier	#LM	Opening	Avg. Opening
2.30	2	2*-18.60	-18.60
2.20	2	2*-18.60	-18.60
2.10	2	2*-18.60	-18.60
2.01	2	2*-18.60	-18.60

2.00	4 **(12)**	4*-18.60	-18.60
1.91	2	2*-18.60	-18.60
1.90	4	4*-18.60	-18.60
1.81	2 **(20)**	2*-18.60	-18.60
1.80	10	10*-18.60	-18.60
1.71	4	4*-18.60	-18.60
1.70	10	10*-18.60	-18.60
1.61	2	2*-18.60	-18.60
1.60	36 **(82)**	34*-18.60, 2*-17.90	-18.56
1.51	10	10*-18.60	-18.60
1.50	30 **(122)**	28*-18.60, 2*-17.90	-18.55

A-Allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
12	25.22/12 2.10	-18.60
24	40.26/20 2.01	-18.60
83	125.92/82 1.54	-18.58
170	186.02/122 1.52	-18.58

C-Allele

Barrier	#LM	Opening	Avg. Opening
2.30	8	-21.30	-21.30
2.20	8 **(16)**	-21.30	-21.30
2.10	8 **(24)**	-21.30	-21.30
2.01	8	-21.30	-21.30
2.00	16	-21.30	-21.30
1.91	8	-21.30	-21.30
1.90	20 **(76)**	-21.30	-21.30
1.81	8	-21.30	-21.30
1.80	92 **(176)**	-21.30	-21.30

C-Allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
16	36/16 2.25	-21.30

24	52.8/24 2.20	-21.30
76	154.16/76 2.03	-21.30
176	334.24/176 1.90	-21.30

Ordered by Opening Energy

A-Allele

Opening	#LM	Barrier	Avg. Barrier
-17.30	5	1.00, 2*0.90, 2*0.80	0.88
-17.90	186	2*1.60, 2*1.50, 6*1.40, 2*1.31, 8*1.30, 2*1.21, 10*1.20, 2*1.11, 14*1.10, 6*1.01, 20*1.00, 2*0.91, 44*0.90, 10*0.81, 26*0.80	2.13

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	1.19	-17.60
20	1.28	-17.90
60	1.20	-17.90
100	1.10	-17.90

C-Allele

Opening	#LM	Barrier	Avg. Barrier
-21.30	14,943	8*2.30, 8*2.20, 8*2.10, 8*2.01, 16*2.00, 8*1.91, 20*2.00, 8*1.81, 92*1.80, 24*1.71, 76*1.70, 16*1.61, 196*1.60, 52*1.51, 132*1.50, 64*1.41, 704*1.40, 104*1.31, 520*1.30, 96*1.21, 600*1.20, 204*1.11, 1696*1.10, 220*1.01, 2204*1.00, 428*0.91, 3092*0.90, 592*0.81, 3748*0.80	1.00

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	2.15	-21.30
20	1.94	-21.30
60	1.80	-21.30
100	1.79	-21.30

10. FGF20 (L = 903)

NM_019851.2 vs. rs12720208
miRNA: miR-433-3p
SNP Pos. 182

	C-allele	U-allele
Target Site:	166 - 187	166 - 187

RNAsubopt and Barrier Setting

Offset	Barrier	C-allele	U-allele	C-allele	U-allele
2.3	0.8	1,213,806	2,514,455	5,937	7,784

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
C-allele	-14.50	-2.356	-9.274	-5.226
U-allele	-12.30	-4.655	-10.123	-2.177

Local Minima

MFE	**C -allele**	**U-allele**
#Pairings	9	10
Opening Energy	-4.10	-7.30
Structure Energy	-159.50	-158.90
MFE Barrier	2.30	2.30
Minima		
Less Pairings	0.6% 34	7.5% 582
Equal Pairings	51.3% 3,048	92.5% 7,201
Greater Pairings	48.1% 2,854	0%
Target Site Identical to MFE	50.9% 3,020	89% 6,927
Minima (- MFE)	**5,936**	**7,783**
Avg. Opening of All Local Minima (excluding MFE)	**-6.40**	**-8.92**

Less/Equal Pairings than MFE

#Pairings	#Minima C-allele	Target Site Approx. Avg.	#Minima U-allele	Target Site Approx. Avg.
6	0.1% 6	-5.80	1.7% 132	-5.80
7	0.5% 28	-6.80	5.8% 450	-6.79
9	51.3% 3,048	-4.10	0	-
10	**30.4% 1,804**	**-7.66**	**92.5% 7,201**	**-9.11**
Total:	**82.3% 4,886**	**-5.43**	**100% 7,783**	**-8.92**
Avg. Barrier:	**0.97**		**1.05**	

Local Minima with Less Pairings

	C-allele	U-allele
#LM Less Pairings:	0.6% 34	7.5% 582
Avg. Target Site Energy	-6.62	-6.57

Less Pairings Avg. Barrier Height	0.88	1.03

Local Minima with Less or Equal Approx. Target Site Energy than the MFE

	C-allele		U-allele	
Less	0.7%	42	11%	856
Equal	50.9%	3,020	44.5%	3,463
Less or Equal	51.6%	3,062	55.5%	4,319

N+ Test

C-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-158.80	70 **(163)**	54*1.60, 1.40, 9*1.20, 6*0.90	1.49	9*-11.10, 3*-10.60, 2*-8.10, 2*-7.30, -4.30, 53*-4.10	-5.49
-158.90	45 **(93)**	36*1.70, 1.60, 5*1.20, 3*0.90	1.59	4*-11.10, 2*-10.60, -8.10, -7.30, 37*-4.10	-5.17
-159.00	23	18*1.80, 3*1.20, 2*0.90	1.64	2*-11.10, -10.60, 20*-4.10	-4.99
-159.10	12 **(25)**	9*1.90, 2*1.20, 0.90	1.70	-11.10, 11*-4.10	-4.68
-159.20	7 **(13)**	6*2.00, 1.20	1.88	7*-4.10	-4.10
-159.30	3	3*2.10	2.10	3*-4.10	-4.10
-159.40	2	2*2.20	2.20	2*-4.10	-4.10
MFE: -159.50	1	2.30	2.30	-4.10	-4.10
-25903.3 Avg. -158.92	**163**	**259.9**	**1.59**	**-841.1**	**-5.16**

C-allele - N+ Averages

Test #LM-MFE	Avg. Structure Energy	Avg. Barrier	Avg. Opening Energy
13	-2070.6/13 -159.28	26.2/13 2.02	-53.3/13 -4.10
25	-3979.8/25 -159.19	46.6/25 1.86	-109.5/25 -4.38
93	-14787.3/93 -159.00	155.9/93 1.68	-457/93 -4.91
163	-25903.3/163 -158.92	259.9/163 1.59	-841.1/163 -5.16

N+ Test

U-allele Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-158.50	51 **(101)**	37*1.80, 4*1.30, 6*1.20, 4*0.90	1.62	25*-11.10, 25*-7.30, -6.80	-9.15
-158.60	28 **(50)**	18*1.90, 2*1.80, 2*1.30, 4*1.20, 2*0.90	1.68	14*-11.10, 14*-7.30	-9.20
-158.70	16 **(22)**	14*2.00, 2*1.20	1.90	8*-11.10, 8*-7.30	-9.20
-158.80	4	4*2.20	2.20	2*-11.10, 2*-7.30	-9.20
MFE: -158.90	2	2*2.30	2.30	-11.10, -7.30	-9.20
-16016.5 Avg. -158.58	**101**	**173.4**	**1.71**	**926.8**	**-9.20**

U-allele - N+ Averages

Test #LM-MFE	Avg. Structure Energy	Avg. Barrier	Avg. Opening Energy
20	-3492.2/22 -158.74	43.8/22 1.99	202.4/22 -9.20
48	-7933/50 -158.66	90.8/50 1.82	460/50 -9.20
101	-16016.5/101 -158.58	173.4/101 1.71	926.8/101 -9.20

Ordered by Deepest Local Minima

C-allele

Barrier	#LM	Opening	Avg. Opening
2.20	2	2*-4.10	-4.10
2.10	3	3*-4.10	-4.10
2.00	6 **(11)**	6*-4.10	-4.10
1.90	9 **(20)**	8*-4.10, 1*-11.10	-4.88
1.80	18	15*-4.10, 1*-10.60, 2*-11.10	-5.24
1.70	35 **(73)**	26*-4.10, 1*-4.30, 1*-8.10, 1*-7.30, 2*-10.60, 4*-11.10	-5.25
1.60	55 **(128)**	38*-4.10, 2*-4.30, 2*-7.30, 2*-8.10, 3*-10.60, 8*-11.10	-5.74

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
11	2.06	-4.10
20	1.99	-4.45
73	1.80	-5.14

128	1.72	-5.40

U-Allele

Barrier	#LM	Opening	Avg. Opening
2.30	1	-11.10	-11.10
2.20	4	2*-11.10, 2*-7.30	-9.20
2.10	6 **(11)**	3*-11.10, 3*-7.30	-9.20
2.00	14 **(25)**	7*-11.10, 7*-7.30	-9.20
1.90	18	9*-11.10, 9*-7.30	-9.20
1.80	39 **(82)**	19*-11.10, 19*-7.30, -6.80	-9.14
1.70	66 **(148)**	32*-11.10, 32*-7.30, 2*-6.80,	-9.13

U-Allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
11	2.15	-9.37
25	51.7 2.07	-9.28
82	156.1 1.90	-9.19
148	1.81	-9.16

Ordered by Opening Energy

C-Allele

Opening	#LM	Barrier	Avg. Barrier
-3.70	14	1.10, 2*1.00, 3*0.90, 8*0.80	0.87
-3.80	28	6*1.00, 8*0.90, 14*0.80	0.87
-4.10	3,020	2*2.20, 3*2.10, 6*2.00, 8*1.90, 15*1.80, 26*1.70, 38*1.60, 53*1.50, 80*1.40, 104*1.30, 302*1.20, 1.14, 278*1.10, 2*1.04, 408*1.00, 3*0.94, 801*0.90, 7*0.84, 883*0.80	1.00

U-Allele

Opening	#LM	Barrier	Avg. Barrier
-2.90	1	0.90	0.90
-5.80	132	1.60, 2*1.50, 3*1.40, 6*1.30, 11*1.20, 18*1.10, 33*1.00, 58*0.90	1.02

11. HOXB5 (L = 952)

NM_002147.3 vs. rs9299
miRNA: miR-7
SNP Pos: 141
Target site: 126 - 154

RNAsubopt and Barrier Setting

Offset	Barrier	G-allele	A-allele	G-allele	A-allele
2.0	0.8	1,383,237	480,436	6,481	3,746

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
G-allele	-21.90	-0.49	-13.61	-8.29
A-allele	-22.40	-1.39	-11.20	-11.20

Local Minima

MFE	G-allele	A-allele
#Pairings	15	15
Opening Energy	-17.60	-18.30
Structure Energy	-302.70	-302.40
MFE Barrier	2.00	2.00
Minima		
Less Pairings	0.2% 16	1.4% 54
Equal Pairings	99.8% 6,464	98.6% 3,691
Greater Pairings	0	0
Identical to MFE	84.9% 5,504	98.6% 3,691
Minima (- MFE)	**6,480**	**3,745**
Avg. Opening of All Local Minima (excluding MFE)	**-17.65**	**-18.21**

#Pairings	#Minima G-allele	Target Site Approx. Avg.	#Minima A-allele	Target Site Approx. Avg.
9	-	-	1.44% 54	-12.20
12	0.2% 16	-13.95	-	-
15	**99.8% 6,464**	**-17.65**	**98.6% 3,691**	**-18.30**
Total Minima	**6,480**		**3,745**	
Avg. Barrier	**1.00**		**0.94**	

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	G-allele	A-allele
Less	4.0% 262	1.4% 54
Equal	84.9% 5,504	98.6% 3,691
Less or Equal	**88.98% 5,766**	**100% 3,745**

N+ Test

G-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-302.30	58 **(109)**	58*1.60	1.60	4*-18.30, 54*-17.60	-17.65
-302.40	22 **(51)**	22*1.70	1.70	4*-18.30, 18*-17.60	-17.73
-302.50	12 **(29)**	12*1.80	1.80	12*-17.60	-17.60
-302.60	12 **(17)**	12*1.90	1.90	12*-17.60	-17.60
-302.70	5	5*2.00	2.00	5*-17.60	-17.60
MFE: -302.70	1	2.00	2.00	-17.60	-17.60

G-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
17	-5144.7/17 -302.63	32.8/17 1.93	299.2/17 -17.60
29	-8774.7/29 -302.58	54.4/29 1.88	-510.4/29 -17.60
51	-15427.5/51 -302.50	91.8/51 1.80	-900.4/61 -17.65
109	-32960.9/109 -302.40	184.6/109 1.69	-1924/109 -17.65

A-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-301.90	36 **(111)**	32*1.50, 2*1.40, 2*1.30	1.48	36*-18.30	-18.30
-302.00	44 **(75)**	38*1.60, 4*1.40, 2*1.30	1.57	44*-18.30	-18.30
-302.10	12 **(31)**	6*1.70, 2*1.40, 4*0.80	1.35	12*-18.30	-18.30
-302.20	12 **(19)**	10*1.80, 2*1.40	1.73	12*-18.30	-18.30
-302.30	4	4*1.90	2.00	4*-18.30	-18.30
-302.40	3	3*2.00	2.00	3*-18.30	-18.30
MFE: -302.40	1	2.00	2.00	-18.30	-18.30

A-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
19	-5742.8/19 -302.25	34.4/19 1.81	347.7/19 -18.30
31	-9368/31	50.6/31	-567.3/31

	-302.19	1.63	-18.30
75	-22656/75 -302.08	119.6/75 1.59	-1372.5/75 -18.30
111	-33524.4/111 -302.02	173/111 1.56	-2031.3/111 -18.30

Ordered by Deepest Local Minima

G-allele

Barrier	#LM	Opening	Avg. Opening
2.00	5	5*-17.60	-17.60
1.90	12	12*-17.60	-17.60
1.80	12	12*-17.60	-17.60
1.70	22	22*-17.60	-17.60
1.60	54	54*-17.60	-17.60

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	19.5/10 1.95	-176/10 -17.60
20	38.2/20 1.91	-352/20 -17.60
60	106.2/60 1.77	-1056/60 -17.60
100	170.2/100 1.70	-1760/100 -17.60

A-allele

Barrier	#LM	Opening	Avg. Opening
2.00	3	3*-18.30	-18.30
1.90	4	4*-18.30	-18.30
1.80	10	10*-18.30	-18.30
1.70	6	6*-18.30	-18.30
1.60	38	38*-18.30	-18.30
1.50	32	32*-18.30	-18.30
1.40	114	114*-18.30	-18.30

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	19/10 1.90	-183/10 -18.30
20	36.7/20 1.84	-366/20 -18.30
60	101/60 1.68	-1098/60 -18.30
100	160.4/100 1.60	-1830/100 -18.30

Ordered by Opening Energy

G-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-13.95	16	2*1.10, 6*1.00, 8*0.90	0.96
-17.00	246	6*1.40, 12*1.30, 12*1.20, 18*1.10, 54*1.00, 102*0.90, 42*0.80	0.97

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	10/10 1.00	-139.5/10 -13.95
20	21/20 1.05	-291.2/20 -14.56
60	69.2/60 1.15	-971.2/60 -16.19
100	109.6/100 1.10	-1651.2/100 -16.51

A-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-12.20	54	6*1.10, 12*1.00, 12*0.90, 24*0.80	0.90
-18.30	3691	3*2.00, 4*1.90, 10*1.80, 6*1.70, 38*1.60, 32*1.50, 114*1.40, 96*1.30, 246*1.20, 222*1.10, 522*1.00, 680*0.90, 1718*0.80	0.94

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	10.6/10 1.06	-122/10 -12.20
20	20.4/20 1.02	-244/20 -12.20
60	60.3/60 1.01	-768.6/60 -12.81
100	127.2/100 1.27	-1500.6/100 -15.01

12. RAD51 (L = 978)

NM_002875.4 vs. rs7180135
miRNA: miR-197-3p
SNP Pos: 718
Target site: 707 - 725

RNAsubopt and Barrier Setting

Offset	Barrier	G-allele	A-allele	G-allele	A-allele
2.2	1.2	9,020,624	11,874,784	3,850	6,291

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
G-allele	-28.60	-0.56	-14.30	-14.31
A-allele	-22.00	-1.20	-15.00	-7.00

Local Minima

MFE	**G-allele**		**A-allele**	
#Pairings	6		6	
Opening Energy	-6.60		-4.70	
Structure Energy	-285.90		-284.00	
MFE Barrier	2.20		2.20	
Minima				
Less Pairings	0.0%	0	0.0%	0
Equal Pairings	100%	3,849	93.9%	5,907
Greater Pairings	0.0%	0	6.1%	383
Identical to MFE	100%	3,849	93.3%	5,867
Minima (- MFE)	**3,849**		**6,290**	
Avg. Opening of All Local Minima (excluding MFE)	**-6.60**		**-4.75**	

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	G-allele		A-allele	
Less	0.0%	0	0.6%	40
Equal	100%	3,849	93.3%	5,867
Less or Equal	**100%**	**3,849**	**93.9%**	**5,907**

N+ Test

G-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-285.40	204 **(376)**	203*1.70, 1.40	1.70	204*-6.60	-6.60
-285.50	134 **(172)**	99*1.80, 32*1.70, 3*1.20	1.76	134*-6.60	-6.60
-285.60	33 **(38)**	24*1.90, 8*1.70, 1.20	1.83	33*-6.60	-6.60

-285.70	2	2.00, 1.80	1.90	2*-6.60	-6.60
-285.90	3	2*2.10, 1.80	2.00	3*-6.60	-6.60
MFE: -285.90	1	2.20	2.20	-6.60	-6.60
-107618.1 Avg. -285.46	**377**	**655.10**	**1.74**	**-2488.19**	**-6.60**

G-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
10	-2857.1/10 -285.71	19.3/10 1.93	66/10 6.60
38	-10853.6/38 -285.62	70.2/38 1.85	-250.8/38 -6.60
172	-49110.6/172 -285.50	306.4/172 1.78	-1135.2/172 -6.60
376	-107332.2/376 -285.50	652.9/376 1.74	-2481.59/376 -6.60

A-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-283.50	207 **(379)**	204*1.70, 1.40, 2*1.20	1.69	207*-4.70	-4.70
-283.60	134 **(172)**	99*1.80, 32*1.70, 3*1.20	1.76	134*-4.70	-4.70
-283.70	33 **(38)**	24*1.90, 8*1.70, 1.20	1.83	33*-4.70	-4.70
-283.80	2	2.00, 1.80	1.90	2*-4.70	-4.70
-283.90	3	2*2.10, 1.80	2.00	3*-4.70	-4.70
MFE: -284.00	1	2.20	2.20	-4.70	-4.70
-107752.3 Avg. -283.60	**380**	**659.20**	**1.73**	**1786.89**	**-4.70**

A-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
10	-2857.1/10 -283.78	19.3/10 1.93	-47/10 -4.70
38	-10781.4/38 -283.70	70.2/38 1.85	-178.6/20 -4.70
172	-48783.79/172 -283.63	306.4/172 1.78	-808.4/172 -4.70
379	-107468.3/379 -283.60	657/379 1.73	-1782.19/379 -4.70

Ordered by Deepest Local Minima

G-allele

Barrier	#LM	Opening	Avg. Opening
2.10	2	2*-6.60	-6.60
2.00	1	-6.60	-6.60
1.90	24	24*-6.60	-6.60
1.80	101	101*-6.60	-6.60

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	19.5/10 1.95	-66/10 -6.60
20	38.5/20 1.93	-132/20 -6.60
60	91.5/60 1.53	-396/60 -6.60
100	183.2/100 1.83	-660/100 -6.60

A-allele

Barrier	#LM	Opening	Avg. Opening
2.10	2	2*-4.70	-4.70
2.00	1	-4.70	-4.70
1.90	24	24*-4.70	-4.70
1.80	101	101*-4.70	-4.70

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	19.5/10 1.95	-47/10 -4.70
20	38.5/20 1.93	-94/20 -4.70
60	91.5/60 1.53	-282/60 -4.70
100	183.2/100 1.83	-470/100 -4.70

Ordered by Opening Energy

G-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-6.60	3,849	2*2.10, 2.00, 24*1.90, 101*1.80, 243*1.70, 238*1.60, 446*1.50, 914*1.40, 1306*1.30, 574*1.20	1.39

G-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	19.5/10 1.95	-66/10 -6.60
20	46.1/20 2.31	-132/20 -6.60
60	111.2/60 1.85	-396/60 -6.60
100	183.2/100 1.83	-660/100 -6.60

A-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-3.70	1	1.20	1.20
-4.00	39	1.50, 3*1.40, 2*1.30, 33*1.20	1.23
-4.70	5867	2*2.10, 2.00, 24*1.90, 101*1.80, 243*1.70, 238*1.60, 446*1.50, 914*1.40, 1306*1.30, 2592*1.20	1.33

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	13.1/10 1.31	-39.7/10 -3.97
20	25.1/20 1.26	-79.7/20 -3.99
60	87.6/60 1.46	-253.7/60 -4.23
100	160.3/100 1.60	-441.7/100 -4.42

13. ORAI1 (L = 1034)

NM_032790.3 vs. rs76753792
miRNA: miR-519a-3p
SNP Pos: 86

	C-allele	U-allele
Target Site:	69 - 88	81 - 102

RNAsubopt and Barrier Setting

Offset	Barrier	C-allele	U-allele	C-allele	U-allele
2.0	0.9	223,532	559,335	577	1,332

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
U-allele	-17.90	-0.02	-23.42	5.62
C-allele	-16.70	-4.32	-9.49	-7.21

Local Minima

MFE	C-allele		U-allele	
#Pairings	6		16	
Opening Energy	-5.40		-27.21	
Structure Energy	-403.50		-405.20	
MFE Barrier	2.00		2.00	
Minima				
Less Pairings	0.0%	0	1.4%	18
Equal Pairings	100%	576	98.6%	1,313
Greater Pairings	0.0%	0	0.0%	0
Identical to MFE	100%	576	49.3%	656
Minima (- MFE)	**576**		**1,331**	
Avg. Opening of All Local Minima (excluding MFE)	**-5.40**		**-27.20**	

#Pairings	#Minima C-allele		Target Site Approx. Avg.	#Minima U-allele		Target Site Approx. Avg.
5	**100%**	**576**	**-5.40**	-		-
...	-		-	-		-
15	-		-	1.4%	18	-26.29
16	-		-	**98.65**	**1,313**	**-27.21**
Total Minima	**100%**	**576**	**-5.40**	**100%**	**1,331**	**-27.20**
Avg. Barrier	**1.16**			**1.12**		

Local Minima with Less Pairings

	C-allele	U-allele
#LM Less Pairings:	0.0%	1.4% 18
Avg. Target Site Energy	-	-26.29
Less Pairings Avg. Barrier Height	-	1.02

Local Minima with Less/Equal Approx. Target Site Energy than the MFE

	C-allele		U-allele	
Less	0.0%	0	1.4%	18
Equal	100%	576	98.6%	1,313
Less or Equal	**100%**	**576**	**100%**	**1,331**

C-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-402.90	35 **(132)**	21*1.40, 10*1.10, 4*1.00	1.27	35*-5.40	-5.40
-403.00	32	18*1.50, 4*1.10, 10*1.00	1.29	36*-5.40	-5.40
-403.10	26 **(65)**	22*1.60, 4*1.00	1.51	26*-5.40	-5.40
-403.20	22 **(39)**	12*1.70, 10*0.90	1.34	22*-5.40	-5.40
-403.30	4	4*1.80	1.80	4*-5.40	-5.40
-403.40	10 **(13)**	10*1.90	1.90	10*-5.40	-5.40
-403.50	3	3*2.00	2.00	3*-5.40	-5.40
MFE: -403.50	1	2.00	2.00	-5.40	-5.40
-53609.7 Avg. -403.10	**133**	**188.6**	**1.42**	**-718.20**	**-5.40**

C-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
39	-15728.1/39 -403.30	61.6/39 1.58	-210.6/20 -5.40
65	-26208.7/65 -403.20	100.8/65 1.55	-351/65 -5.40
132	-53206.2/132 -403.10	186.6/132 1.41	-712.8/100 -5.40

U-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-404.80	52 **(111)**	44*1.60, 8*1.00	1.51	52*-27.21	-27.21
-404.90	24 **(59)**	24*1.70	1.70	24*-27.21	-27.21
-405.00	8	8*1.80	1.80	8*-27.21	-27.21
-405.10	20 **(27)**	20*1.90	1.90	20*-27.21	-27.21
-405.20	7	7*2.00	2.00	7*-27.21	-27.21
MFE: -405.20	1	2.00	2.00	-27.21	-27.21
-45350.8 Avg. -404.92	**112**	**187.6**	**1.68**	**-3047.52**	**-27.21**

U-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
111	-44945.6/111 -404.90	185.6/111 1.67	-3020.31/111 -27.21
59	-23896/59 -405.00	107.2/59 1.82	-1605.39/59 -27.21
27	-10938.4/27 -405.10	52/27 1.93	-737.67/20 -27.21

Ordered by Deepest Local Minima

C-allele

Barrier	#LM	Opening	Avg. Opening
2.00	3	3*-5.40	-5.40
1.90	10	10*-5.40	-5.40
1.80	4	4*-5.40	-5.40
1.70	12	12*-5.40	-5.40
1.60	22	22*-5.40	-5.40
1.50	18	18*-5.40	-5.40
1.40	21	21*-5.40	-5.40
1.30	38	38*-5.40	-5.40

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
13	25/13 1.92	-70.2/13 -5.40
29	52.6/29 1.81	-156.6/29 -5.40
69	114.8/69 1.66	-372.6/69 -5.40
128	193.6/128	-691.19/100

	1.51	-5.40

U-allele

Barrier	#LM	Opening	Avg. Opening
2.00	7	7*-27.21	-27.21
1.90	20	20*-27.21	-27.21
1.80	8	8*-27.21	-27.21
1.70	24	24*-27.21	-27.21
1.60	44	44*-27.21	-27.21

U-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
7	14/7 2.00	-190.47/10 -27.21
27	52/27 1.93	-734.67/27 -27.21
59	107.2/59 1.82	-1632.6/59 -27.21
103	177.6/100 1.72	-2802.63/100 -27.21

Ordered by Opening Energy

C-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-5.40	576	3*2.00, 10*1.90, 4*1.80, 12*1.70, 22*1.60, 18*1.50, 21*1.40, 38*1.30, 4*1.20, 131*1.10, 181*1.00, 83*1.20, 49*0.90	1.16

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	19.3/10 1.93	-54/10 -5.40
20	37.3/20 1.87	-108/20 -5.40
60	101.3/60 1.69	-324/60 -5.40
100	157.2/100 1.57	-540/100 -5.40

U-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-26.31	18	4*1.10, 14*1.00	1.02
-27.21	1313	7*2.00, 20*1.90, 8*1.80, 24*1.70, 44*1.60, 36*1.50, 42*1.40, 76*1.30, 174*1.20, 242*1.10, 382*1.00, 258*0.90	1.13

U-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	10.4/10 1.04	-263.1/10 -26.31
20	22.4/20 1.12	-528/20 -26.40
60	96.7/60 1.61	-1616.4/60 -26.94
100	162.4/100 1.62	-2704.8/100 -27.05

14. RAP1 (L = 1078)

NM_02884.2 vs. rs6573
miRNA: miR-196a
SNP Pos: 366
Target Site: 348 - 370

RNAsubopt and Barrier Setting

Offset	Barrier	C-allele	A-allele	C-allele	A-allele
2.0	0.9	1,689,428	1,689,428	238	238

STarMir Target Site Energies

	dG_hybrid	dG_nucl	dG_open	dG_total
C-allele	-16.70	-3.97	-5.40	-11.30
A-allele	-21.30	-6.97	-5.05	-16.25

Local Minima

MFE	**C-allele**	**A-allele**
#Pairings	12	12
Opening Energy	-5.10	-5.10
Structure Energy	-197.80	-197.80
MFE Barrier	2.00	2.00
Minima		
Less Pairings	0.0%	0.0%
Equal Pairings	100% 237	100% 237
Greater Pairings	0.0%	0.0%
Identical to MFE	100% 237	100% 237
Minima (- MFE)	**237**	**237**
Avg. Opening of All Local Minima (excluding MFE)	**-5.10**	**-5.10**

C-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-196.90	46 **(143)**	46*1.10	1.10	46*-5.10	-5.10
-197.00	30	30*1.20	1.20	30*-5.10	-5.10
-197.10	28 **(67)**	22*1.30, 6*0.90	1.21	28*-5.10	-5.10
-197.20	12	10*1.40, 2*1.30	1.38	12*-5.10	-5.10
-197.30	10 **(27)**	10*1.50	1.50	10*-5.10	-5.10
-197.40	8	6*1.60, 2*1.50	1.58	8*-5.10	-5.10
-197.50	6	6*1.70	1.70	6*-5.10	-5.10
-197.70	2	2*1.90	1.90	2*-5.10	-5.10
-197.80	1	2.00	2.00	-5.10	-5.10
MFE: -197.80	1	2.00	2.00	-5.10	-5.10

-28380.8 Avg. -197.09	**144**	**182.8**	**1.27**	**-734.4**	**-5.10**

C-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
143	-28183/143 -197.08	180.8/143 1.26	-729.3/143 -5.10
67	-13215.6/67 -197.25	94.2/67 1.41	-341.7/67 -5.10
27	-5330.4/27 -197.42	43.6/27 1.61	-137.7/27 -5.10

A-allele - Local Minima Distribution

E	#LM	Barrier	Avg. Barrier	Opening	Avg. Opening
-196.90	46 **(143)**	46*1.10	1.10	46*-5.10	-5.10
-197.00	30	30*1.20	1.20	30*-5.10	-5.10
-197.10	28 **(67)**	22*1.30, 6*0.90	1.21	28*-5.10	-5.10
-197.20	12	10*1.40, 2*1.30	1.38	12*-5.10	-5.10
-197.30	10 **(27)**	10*1.50	1.50	10*-5.10	-5.10
-197.40	8	6*1.60, 2*1.50	1.58	8*-5.10	-5.10
-197.50	6	6*1.70	1.70	6*-5.10	-5.10
-197.70	2	2*1.90	1.90	2*-5.10	-5.10
-197.80	1	2.00	2.00	-5.10	-5.10
MFE: -197.80	1	2.00	2.00	-5.10	-5.10
-28380.8 Avg. -197.09	**144**	**182.8**	**1.27**	**-734.4**	**-5.10**

A-allele - N+ Averages

Test #LM-MFE	Structure Energy	Barrier	Opening Energy
27	-5330.4/27 -197.42	43.6/27 1.61	-137.7/27 -5.10
67	-13215.6/67 -197.25	94.2/67 1.41	-341.7/67 -5.10
143	-28183/143 -197.08	180.8/143 1.26	-729.3/143 -5.10

Ordered by Deepest Local Minima

C-allele

Barrier	#LM	Opening	Avg. Opening
2.00	1	-5.10	-5.10

1.90	2	2*-5.10	-5.10
1.70	6	6*-5.10	-5.10
1.60	6 **(15)**	6*-5.10	-5.10
1.50	12 **(27)**	12*-5.10	-5.10
1.40	10	10*-5.10	-5.10
1.30	24 **(61)**	24*-5.10	-5.10
1.20	30	30*-5.10	-5.10
1.10	46 **(137)**	46*-5.10	-5.10

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
15	25.6/15 1.71	-76.5/15 -5.10
27	43.6/27 1.61	-137.7/27 -5.10
61	88.1/61 1.46	-311.1/61 -5.10
137	175.4/128 1.28	-698.7/137 -5.10

A-allele

Barrier	#LM	Opening	Avg. Opening
2.00	1	-5.10	-5.10
1.90	2	2*-5.10	-5.10
1.70	6	6*-5.10	-5.10
1.60	6	6*-5.10	-5.10
1.50	12	12*-5.10	-5.10
1.40	10	10*-5.10	-5.10
1.30	24	24*-5.10	-5.10
1.20	30	30*-5.10	-5.10
1.10	46	46*-5.10	-5.10

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
15	25.6/15 1.71	-76.5/15 -5.10
27	43.6/27 1.61	-137.7/27 -5.10
61	88.1/61 1.46	-311.1/61 -5.10
137	175.4/128 1.28	-698.7/137 -5.10

Ordered by Opening Energy

C-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-5.10	237	2.00, 2*1.90, 6*1.70, 6*1.60, 12*1.50, 10*1.40, 24*1.30, 30*1.20, 46*1.10, 77*1.00, 22*0.90	1.45

C-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	17.6/10 1.76	-51/10 -5.10
20	33.1/20 1.66	-102/20 -5.10
60	87.5/60 1.46	-306/60 -5.10
100	134.7/100 1.35	-510/100 -5.10

A-allele - Opening Energy

Opening	#LM	Barrier	Avg. Barrier
-5.10	237	2.00, 2*1.90, 6*1.70, 6*1.60, 12*1.50, 10*1.40, 24*1.30, 30*1.20, 46*1.10, 77*1.00, 22*0.90	1.45

A-allele - Averages

Test #LM-MFE	Avg. Barrier	Avg. Opening Energy
10	17.6/10 1.76	-51/10 -5.10
20	33.1/20 1.66	-102/20 -5.10
60	87.5/60 1.46	-306/60 -5.10
100	134.7/100 1.35	-510/100 -5.10

.5 microRNA Target Prediction Based upon Metastable RNA Secondary Structures - RNAStrucTar predictions

LIG3/mir-221

Wild type A allele is expected to be weaker when binding with microRNA-124.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-17.69	-4.59	-20.22	-7.12	-84.789	-5.98
Variant	-29.19	-8.76	-35.13	-13.28	-99.072	-7.45
Δ (stronger – weaker)	-11.5	-4.17	-14.91	-6.16	-14.283	-1.47

MSLN/mir611

Wild type G allele is expected to be stronger when binding with microRNA-611.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-6.61	-5.45 (opening -20.9)	-8.64	-8.64 (opening -16.61)	-121.0	-8.47
Variant	-4.39	-1.37 (opening -17.96)	-4.74	-4.74 (opening -14)	-145.9	-7.84
Δ (stronger – weaker)	-2.22	-4.08	-3.9	-3.9	24.9	-0.63

CGA/mir1302

Wild type T allele is expected to be weaker when binding with microRNA-1302.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-12.10	-4.37 (opening -20.01)	-22.85	-7.62 (opening -12.8)	-23.57	-7.87
Variant	-17.65	-6.36 (opening -19.71)	-28.88	-9.63 (opening -10.65)	-27.61	-8.99
Δ (stronger – weaker)	-5.55	-1.99	-6.03	-2.01	-4.04	-1.12

CBR1/mir-574

Wild type G allele is expected to be weaker when binding with microRNA-284.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-31.95	-12.96	-40.75	-16.88	-156.19	-17.94
Variant	-37.96	-16.06	-42.61	-17.08	-176.74	-19.21
Δ (stronger – weaker)	-6.01	-3.1	-1.86	-0.2	-20.55	-1.27

HTR3E/mir-510

Wild type G allele is expected to be stronger when binding with microRNA-302.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-13.45	-1.71	-27.67	-11.9	-83.24	-16.41
Variant	-19.74	-3.40	-30.26	-10.54	-80.94	-12.28
Δ (stronger – weaker)	6.29	1.69	2.59	-1.36	-2.31	-4.13

KRT81/mir17b

Wild Type G allele is expected to be the stronger allele and the weaker allele is the variant C allele.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-13.92	-10.74	-16.21	-13.11	-45.69	-7.25
Variant	-3.58	NA	NA	NA	-46.26	-5.65
Δ (stronger – weaker)	-10.34	NA	NA	NA	0.57	-1.6

KRT81/mir-20b

Wild Type G allele is expected to be the stronger allele and the weaker allele is the variant C allele.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-16.32	-13.34	-18.47	-15.5	-36.72	-8.14
Variant	-10.20	-9.02	-11.77	-8.8	-36.62	-7.45
Δ (stronger – weaker)	-6.12	-4.32	-6.7	-6.7	-0.1	-0.69

SPI1/mir-569 /369

Wild type C allele is expected to be stronger when binding with microRNA-569.

Wild Type: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-16.56	NA	NA	NA	-26.43	-5.75
Variant	-16.56	NA	NA	NA	-25.32	-3.98
Δ (stronger – weaker)	0	NA	NA	NA	-1.11	-1.77

HLA_G/mir-148a

Wild type C allele is expected to be weaker when binding with microRNA-148a.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-21.62	-10.47	-28.94	-14.56	-41.40	-1.16
Variant	-41.22	-15.44	-50.50	-20.40	-53.22	-6.49
Δ (stronger – weaker)	-19.6	-4.97	-21.56	-5.84	-11.82	-5.33

MTHFD1/mir-197

Wild type A allele is expected to be weaker when binding with microRNA-197.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-16.58	+	-16.58	+	-81.87	-8.43
Variant	-23.88	-3.5	-23.88	-3.5	-80.44	-11
Δ (stronger – weaker)	-7.3	-3.6+	-7.3	-3.6+	1.43	-2.57

NFKBIA/mir-449a

Wild type G allele is expected to be the weaker allele while the variant A allele to be the stronger.

Wild Type: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-7.45	NA	NA	NA	-61.21	-9.73
Variant	-12.15	NA	NA	NA	-83.11	-10.17
Δ (stronger – weaker)	-4.7	NA	NA	NA	-21.90	-0.44

TYMS/mir-1248

Wild type A allele is expected to be weaker when binding with microRNA-1248.

Wild Type: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-10.10	NA	NA	NA	-113.72	-8.51
Variant	-10.30	+ (opening -16.79)	-10.30	+ (opening -16.79)	-112.08	-8.43
Δ (stronger – weaker)	-0.20	NA	NA	NA	1.64	0.08

CCNE1/mir-151a

Wild type C allele is expected to be stronger when binding with microRNA-151a.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-9.22	+(opening -58.64)	-10.1	+(opening -57.46)	-87.15	-3.64
Variant	-9.22	+(opening -53.64)	-10.1	+(opening -52.46)	-86.77	-3.64
Δ (stronger – weaker)	0	0	0	0	-0.38	0

NPM1-mir337

The variant delT allele is expected to have stronger binding with miR-337.

Wild type: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-10.6	NA	NA	NA	-90.68	-3.72
Variant	-32.9	-10.6	-32.9	-10.6	-96.49	-4.87
Δ (stronger – weaker)	-22.3	-10.6+	NA	NA	-5.81	-1.15

AGT-mir31

The wild type C allele is expected to be the stronger allele while the A allele is the weaker.

Wild Type: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-8.2	NA	NA	NA	-12.62	-5.17
Variant	-8.2	NA	NA	NA	-12.68	-5.16
Δ (stronger – weaker)	0	0	0	0	0.06	-0.01

AGT-mir584

The wild type C allele is expected to be the stronger allele while the A allele is the weaker.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-6.1	+	-8.38	-2.85	-129.89	-6.22
Variant	-6.1	NA	NA	NA	-129.68	-2.87
Δ (stronger – weaker)	0	NA	NA	NA	-0.21	-3.35

NCSTN-mir455

Wild type allele is expected to be stronger when binding with microRNA-455.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	+	+(opening -18.12)	+	+ (opening -17.40)	-18.23	-7.59
Variant	+	+ (opening -44.34)	+	+ (opening -44.34)	-10.40	-2.73
Δ (stronger – weaker)	0	0	0	0	-7.83	-4.86

SMUG1-mir770

The variant T allele is expected to be the stronger allele. The G allele is the wild type.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-14.1	+	-14.1	+	-231.58	-12.66
Variant	-19.49	-4.7	-20.9	-4.7	-233.88	-13.25
Δ (stronger – weaker)	-5.39	-4.7+	-6.8	-4.7+	-2.3	-0.59

PARP1/mir-145a

Wild type T allele is expected to be stronger when binding with microRNA-145a.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-25.2	2.94	-38.46	-7.8	-137.71	-2.53
Variant	-22.84	12.68	-24.20	11.8	-103.37	-2.51
Δ (stronger – weaker)	-2.36	-9.74	-14.26	-7.8+	-34.34	-0.02

WFS1/mir-668

Wild type G allele is expected to be stronger when binding with microRNA-668.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-26.4	-4.3	-26.4	-4.3	-221.57	-16.47
Variant	-18.1	NA	NA	NA	-200.81	-12.13
Δ (stronger – weaker)	-8.3	NA	NA	NA	-20.76	-4.34

MSX1/mir-3649

Wild type A allele is expected to be stronger when binding with microRNA-3649.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-40.6	-5.95 (opening -18.7)	-40.6	-5.95 (opening -18.7)	-116.17	-12.02
Variant	-28.4	-2.85 (opening -24.9)	-28.4	-2.85 (opening -24.9)	-122.80	-10.32
Δ (stronger – weaker)	-12.2	-3.1	-12.2	-3.1	6.63	-1.7

EFNA1/mir-200c

Wild type G allele is expected to be stronger when binding with microRNA-200c.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-13.38	-4.44	-16.6	-6	-86.29	-6.28
Variant	-11.04	NA	NA	NA	-99.35	-2.93
Δ (stronger – weaker)	-2.34	NA	NA	NA	13.06	-3.35

IL23R/mir-7e

Wild type C allele is expected to be stronger when binding with microRNA-72.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-36.5	+	-36.5	+	-185.39	-14.23
Variant	-36.5	+	-36.5	+	-186.15	-12.11
Δ (stronger – weaker)	0	0	0	0	0.76	-2.12

RYR3/mir-367

Wild type A allele is expected to be stronger when binding with microRNA-367.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-2.4	+	-2.1	+	-43.42	-6.97
Variant	-1.4	+	-1.4	+	-32.60	-6.97
Δ (stronger – weaker)	-1	0	-0.7	0	-10.82	0

AGTR1/mir-155

Wild type A allele is expected to be stronger when binding with microRNA-155.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-20.5	+	-20.5	+	-81.05	-5.40
Variant	-16.20	NA	NA	NA	-63.47	-5.19
Δ (stronger – weaker)	-4.3	NA	NA	NA	-17.58	-0.21

FGF20/mir-433

Wild type C allele is expected to be stronger when binding with microRNA-433.

Wild Type: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-21.44	NA	NA	NA	-304.83	-11.08
Variant	-21.62	NA	NA	NA	-313.92	-11.08
Δ (stronger – weaker)	0.18	NA	NA	NA	9.10	0

DROSHA/mir-27b

Wild type T allele is expected to be stronger when binding with microRNA-27b.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-16.8	+ (opening -20.5)	-20.4	+ (opening -17.15)	-28.56	-6.10
Variant	-16.8	+ (opening -20.5)	-19.08	+ (opening -17.13)	-38.69	-6.01
Δ (stronger – weaker)	0	0	-1.32	0	10.13	-0.09

HOXB5/mir-7

Wild type G allele is expected to be weaker when binding with microRNA-7.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-6.8	+	-6.8	+	-96.95	-8.39
Variant	-26.22	+	-26.22	+	-95.20	-10.25
Δ (stronger – weaker)	-19.42	0	-19.42	0	1.751	-1.86

RAD51/mir-197

Wild type G allele is expected to be stronger when binding with microRNA-197.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-39.7	-15.9	-39.7	-15.9	-58.205	-12.50
Variant	-30.1	-11.1	-30.1	-11.1	-70.83	-12.48
Δ (stronger – weaker)	-9.6	-4.8	-9.6	-4.8	12.625	-0.02

REV3L/mir-25

Wild type C allele is expected to be weaker when binding with microRNA-25.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-37.52	-3.3	-45.74	-4.6	-72.74	-8.24
Variant	-37.98	-6.14	-49.62	-9.02	-79.16	-8.26
Δ (stronger – weaker)	-0.46	-2.84	-3.88	-4.42	-6.427	0

ORAI1/mir-519a

Wild type C allele is expected to be stronger when binding with microRNA-519a

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-0.3	+	-0.3	+	-58.00	-10.66
Variant	-0.3	+	-0.3	+	-50.44	-10.66
Δ (stronger – weaker)	0	0	0	0	-7.558	0

RAP1A/mir-196a

Wild type C allele is expected to be weaker when binding with microRNA-196a.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-16.2	-7.6	-16.2	-7.6	-112.26	-9.02
Variant	-29.5	-12.3	-29.5	-12.3	-131.05	-12.08
Δ (stronger – weaker)	-13.3	-4.7	-13.3	-4.7	-18.79	-3.06

APP/mir-147

Wild type T allele is expected to be stronger when binding with microRNA-147.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-56.68	+	-56.68	+	-141.242	-12.63
Variant	-45.45	+	-45.45	+	-134.564	-11.45
Δ (stronger – weaker)	-11.23	0	-11.23	0	-6.678	-1.18

IL1A/mir-122

The Variant TTCA insertion is expected be the weaker allele.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-11.01	+	-11.01	+	-269.45	-11.81
Variant	-11.01	+	-11.01	+	-266.21	-11.07
Δ (stronger – weaker)	0	0	0	0	-3.24	-0.74

CD133/mir-135b

Wild type A allele is expected to be weaker when binding with microRNA-135b.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-5.36	+	-5.36	+	-89.73	-6.97
Variant	-9.26	-2.1	-9.26	-2.1	-101.29	-10.22
Δ (stronger – weaker)	-3.9	-2.1+	-3.9	-2.1+	-11.56	-3.25

PDCD1/mir-4717

Wild type G allele is expected to be stronger when binding with microRNA-4717.

Wild Type: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

Variant: no sites with SNP position inside, therefore no S_u for Case A and no Case B.

	Case A		Case B		STarMir Linear Sum	PITA score
	S_{sum}	S_u	S_{sum}	S_u		
Wild Type	-16.6	NA	NA	NA	-486.86	-11.85
Variant	-14.3	NA	NA	NA	-487	-10.33
Δ (stronger – weaker)	-2.3	NA	NA	NA	0.14	-1.52

Bibliography

[1] the International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001. 1

[2] ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.)*, 306(5696):636–640, 2004. 1

[3] R. Sachidanandam, D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822):928–933, 2001. 1

[4] International HapMap Consortium. The International HapMap Project. *Nature*, 426(6968):789–796, 2003. 1, 36

[5] Ignacio Tinoco Jr and Carlos Bustamante. How {RNA} folds. *Journal of Molecular Biology*, 293(2):271–281, 1999. 3

[6] Kvin Darty, Alain Denise, and Yann Ponty. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15):1974–1975, 2009. 3

[7] Mariusz Popenda, Marta Szachniuk, Maciej Antczak, Katarzyna J. Purzycka, Piotr Lukasiak *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Research*, 40(14):e112, 2012. 3

[8] Chen, Yu and Varani, Gabriele. RNA Structure. *eLS*, 2010. 3

[9] Hendrix, D.K., Brenner, S.E. and Holbrook, S.R. RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics*, 38(3):221–243, 2005. 3

[10] Lorenz, Ronny and Hofacker, Ivo L. and Stadler, Peter F. RNA folding with hard and soft constraints. *Algorithms for Molecular Biology*, 11(1):8, 2016. 5

[11] Ivo L. Hofacker. Vienna RNA secondary structure server. *Nucleic Acids Research*, 31(13):3429–3431, 2003. 6, 7, 9, 24, 25, 26

[12] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, 2003. 6, 26, 99

[13] Jaces R. Fresco, B. M. Alberts, and P. Doty. Some molecular details of the secondary structure of ribonucleic acid. *Nature*, 188:98–101, 1960. 6

[14] Charles Delisi and Donald M. Crothers. Prediction of RNA Secondary Structure. *Proceedings of the National Academy of Sciences of the United States of America*, 68(11):2682–2685, 1971. 6

[15] I. Tinoco, O. C. Uhlenbeck, and M. D. Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, 230(5293):362–367, 1971. 6

[16] Tumanyan VG, Sotnikova LE, Kholopov AE. On the identification of secondary RNA structure from the nucleotide sequence. *Doklady Akademii Nauk SSSR*, 166:1465–1468, 1966 6

[17] R. Bellman. On the Theory of Dynamic Programming. *Proceedings of the National Academy of Sciences of the United States of America*, 38(8):716–719, 1952. 6

[18] J. M. Pipas and J. E. McMahon. Method for predicting RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, 72(6):2017–2021, 1975. 6

[19] M. S. Waterman and T. F. Smith. RNA secondary structure: A complete mathematical analysis. *Mathematical Biosciences*, 42:257–266, 1978. 6

[20] Michael Zuker and Patrick Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1):133–148, 1981. 7

[21] Sean R Eddy. How do RNA folding algorithms work. *Nature Biotechnology*, 22:1457–1458, 2004. 7

[22] I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie / Chemical Monthly*, 125(2):167–188, 1994. 7, 21, 27

[23] M. Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244(4900):48–52, 1989. 7

[24] Kertesz,M. *et al.* The role of site accessibility in microRNA target recognition. *Nature Genetics*, 39:1278–84, 2007. 7, 21, 27, 81, 103, 104, 105, 110

[25] Tianbing Xia, Jr. John SantaLucia, Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37(42):14719–14735, 1998. 7

[26] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19):7287–7292, 2004. 7, 78, 103

[27] Rune B. Lyngso and Christian N. S. Pedersen. RNA Pseudoknot Prediction in Energy-Based Models. *Journal of Computational Biology*, 7(3-4):409–427, 2000. 7

[28] Christoph Flamm and Ivo L. Hofacker. Beyond energy minimization: approaches to the kinetic folding of RNA. *Monatshefte für Chemie - Chemical Monthly*, 139(4):447–457, 2008. 8

[29] Peter Schuster. Prediction of RNA secondary structures: from theory to models and real molecules. *Reports on Progress in Physics*, 69(5):1419, 2006. 8

[30] C Flamm, W Fontana, I L Hofacker, and P Schuster. RNA folding at elementary step resolution. *RNA*, 6(3):325–338, 2000. 8

[31] Dietmar Porschke. Model calculations on the kinetics of oligonucleotide double helix coil transitions. Evidence for a fast chain sliding reaction. *Biophysical Chemistry*, 2(2):83–96, 1974. 8

[32] Michael Geis, Christoph Flamm, Michael T. Wolfinger, Andrea Tanzer, Ivo L. Hofacker *et al.* Folding kinetics of large RNAs. *Journal of Molecular Biology*, 379(1):160–173, 2008. 8

[33] Steven R Morgan and Paul G Higgs. Barrier heights between ground states in a model of rna secondary structure. *Journal of Physics A: Mathematical and General*, 31(14):3153, 1998. 8

[34] E. M. Mahen, P. Y. Watson, J. W. Cottrell, and M. J. Fedor. mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLOS Biology*, 8(2):1–14, 2010. 8

[35] Ivo L. Hofacker, Christoph Flamm, Christian Heine, Michael T. Wolfinger, Gerik Scheuermann, and Peter F. Stadler. BarMap: RNA folding on dynamic energy landscapes. *RNA*, 16(7):1308–1316, 2010. 8

[36] Solomatin,S. *et al.* Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, 463:681–684, 2010. 8, 41, 77

[37] Wuchty,S. *et al.* Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49(2):145–165, 1999. 9, 77, 101

[38] Michael S. Waterman and Thomas H. Byers. A dynamic programming algorithm to find all solutions in a neighborhood of the optimum. *Mathematical Biosciences*, 77(1-2):179–188, 1985. 10

[39] C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger. Barrier Trees of Degenerate Landscapes, *Zeitschrift für Physikalische Chemie*, 216(2):155–173, 2002. 10, 101

[40] Lorenz,W.A., Clote,P. Computing the partition function for kinetically trapped RNA secondary structures. *PLoS One*, 6(1):1–17, 2011. 10, 77

[41] Jiabin Huang, Rolf Backofen, and Björn Voß. Abstract folding space analysis based on helices. *RNA*, 18(12):2135–2147, 2012. 10

[42] Marcel Kucharìk, Ivo L. Hofacker, Peter F. Stadler, and Jing Qin. Basin Hopping Graph: a computational framework to characterize RNA folding landscapes. *Bioinformatics*, 30(14):2009–2017, 2014. 10, 106

[43] Yuan Li and Shaojie Zhang. Finding stable local optimal RNA secondary structures. *Bioinformatics*, 27(21):2994–3001, 2011. 10

[44] Saffarian,A. *et al.* RNA locally optimal secondary structures. *Journal of Computational Biology*, 19(10):1120–1133, 2012. 10, 77

[45] Sean R Eddy. Noncoding RNA genes and the modern RNA world. *Nature Reviews Genetics*, 2:919–929, 2001. 10

[46] John S. Mattick and Igor V. Makunin. Non-coding RNA. *Human Molecular Genetics*, 15(suppl 1):R17–R29, 2006. 12

[47] Thomas R. Cech and Joan A. Steitz. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, 157(1):77–94, 2014. 12

[48] Ha, Minju and Kim, V. Narry. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, 15:509-524, 2014. 12, 14

[49] Antony Rodriguez, Sam Griffiths-Jones, Jennifer L. Ashurst, and Allan Bradley. Identification of mammalian microRNA host genes and transcription units. *Genome Research*, 14(10a):1902–1910, 2004. 12

[50] Rosalind C. Lee, Rhonda L. Feinbaum, and Victor Ambros. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854, 1993. 12

[51] Bruce Wightman, Ilho Ha, and Gary Ruvkun. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75(5):855–862, 1993. 12

[52] Amy E. Pasquinelli, Brenda J. Reinhart, Frank Slack, Mark Q. Martindale, Mitzi I. Kuroda *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–89, 2000. 12

[53] B. J. Reinhart, F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403(6772):901–906, 2000. 12

[54] Yoontae Lee, Minju Kim, Jinju Han, Kyu-Hyun H. Yeom, Sanghyuk Lee *et al.* MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, 23(20):4051–4060, 2004. 12

[55] Glen M. Borchert, William Lanier, and Beverly L. Davidson. RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology*, 13(12):1097–1101, 2006. 12

[56] Yoontae Lee, Chiyoung Ahn, Jinju Han, Hyounjeong Choi, Jaekwang Kim *et al.*The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956):415–419, 2003. 12

[57] Emily Bernstein, Amy A. Caudy, Scott M. Hammond, and Gregory J. Hannon. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409:363–366, 2001. 12

[58] D. P. Bartel. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297, 2004. 14, 40

[59] Choo, Kong Bung and Soon, Yuen Loon and Nguyen, Phan Nguyen Nhi and Hiew, Michele Sook Yuin and Huang, Chiu-Jung. MicroRNA-5p and -3p co-expression and cross-targeting in colon cancer cells. *Journal of Biomedical Science*, 21(1):95, 2014. 14

[60] Yang, Xiangling and Du, William W. and Li, Haoran and Liu, Fengqiong and Khorshidi, Anna and Rutnam, Zina Jeyapalan and Yang, Burton B. Both mature miR-17-5p and passenger strand miR-17-3p target TIMP3 and induce prostate tumor growth and invasion. *Nucleic Acids Research*, 41(21):9688–9704, 2013. 14

[61] Gunter Meister, Markus Landthaler, Agnieszka Patkaniowska, Yair Dorsett, Grace Teng, and Thomas Tuschl. Human argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, 15(2):185–197, 2004. 14

[62] Julia Winter, Stephanie Jung, Sarina Keller, Richard I. Gregory, and Sven Diederichs. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology*, 11(3):228–234, 2009. 14

[63] Huili Guo, Nicholas T. Ingolia, Jonathan S. Weissman and David P. Bartel Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466:835-840, 2010. 14

[64] Robin C. Friedman, Kyle Kai-How Farh, Christopher B. Burge, and David P. Bartel. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105, 2009. 14, 20, 25

[65] Ana Kozomara and Sam Griffiths-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42(D1):D68–D73, 2014. 14, 32, 40

[66] Marc Friedlander, Esther Lizano, Anna Houben, Daniela Bezdan, Monica Banez-Coronel *et al.* Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biology*, 15(4):R57, 2014. 14

[67] Yu Li and Kris V. Kowdley. MicroRNAs in common human diseases. *Genomics, Proteomics & Bioinformatics*, 10(5):246–253, 2012. 15

[68] Yang Li, Chengxiang Qiu, Jian Tu, Bin Geng, Jichun Yang *et al.* HMDD v2.0: a database for experimentally supported human microRNA and dis-

ease associations. *Nucleic Acids Research*, 42(D1):D1070–D1074, 2014. 15, 16

[69] Lu, Jun and Getz, Gad and Miska, Eric A. and Alvarez-Saavedra *et al.* MicroRNA expression profiles classify human cancers. *Nature*, 435:834-838, 2005. 15

[70] Ryan M. O'Connell, Dinesh S. Rao, Aadel A. Chaudhuri, and David Baltimore. Physiological and pathological roles for microRNAs in the immune system. *Nature Reviews Immunology*, 10(2):111–122, 2010. 15

[71] Ramiro Garzon, George A. Calin, and Carlo M. Croce. MicroRNAs in cancer. *Annual Review of Medicine*, 60(1):167–179, 2009. 15

[72] Zvi Bentwich. MicroRNAs play a central role during cancer and viral infections and can serve as targets for therapy. *Retrovirology*, 3(Suppl 1), 2006. 15

[73] Isaac S. Chan and Geoffrey S. Ginsburg. Personalized medicine: Progress and promise. *Annual Review of Genomics and Human Genetics*, 12(1):217–244, 2011. 16

[74] Qinghua Jiang, Yadong Wang, Yangyang Hao, Liran Juan, Mingxiang Teng *et al.* miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research*, 37(suppl 1):D98–D104, 2009. 16

[75] Donald E. Kuhn, Mickey M. Martin, David S. Feldman, Alvin V. Terry Jr., Gerard J. Nuovo, and Terry S. Elton. Experimental validation of miRNA targets. *Methods*, 44(1):47–54, 2008. 16

[76] Daniel W. Thomson, Cameron P. Bracken, and Gregory J. Goodall. Experimental strategies for microRNA target identification. *Nucleic Acids Research*, 39(16):6845–6853, 2011. 16, 17

[77] Brian Parker and Jiayu Wen. Predicting microRNA targets in time-series microarray experiments via functional data analysis. *BMC Bioinformatics*, 10(Suppl 1):S32, 2009. 17

[78] Marshall Thomas, Judy Lieberman, and Ashish Lal. Desperately seeking microRNA targets. *Nature Structural & Molecular Biology*, 17(10):1169–1174, 2010. 17

[79] Matthias Selbach, Björn Schwanhäusser, Nadine Thierfelder, Zhuo Fang, Raya Khanin, and Nikolaus Rajewsky. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63, 2008. 17, 18, 34

[80] Lee P. Lim, Nelson C. Lau, Philip Garrett-Engele, Andrew Grimson, Janell M. Schelter *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433(7027):769–773, 2005. 17

[81] Daehyun Baek, Judit Villén, Chanseok Shin, Fernando D. Camargo, Steven P. Gygi and David P. Bartel. The impact of microRNAs on protein output. *Nature*, 455(7209):64–71, 2008. 18

[82] Matthew W. Jones-Rhoades and David P. Bartel. Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Molecular Cell*, 14(6):787–799, 2004. 18

[83] Julius Brennecke, Alexander Stark, Robert B Russell, and Stephen M Cohen. Principles of microRNA-target recognition. *PLOS Biology*, 3(3):e85, 2005. 18

[84] Can Alkan, Emre Karakoç, Joseph H. Nadeau, S. Cenk Sahinalp, and Kaizhong Zhang. RNA-RNA Interaction Prediction and Antisense RNA Target Search. *Journal of Computational Biology*, 13(2):267–282, 2006. 18

[85] B. P. Lewis, I. hung Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge. Prediction of mammalian microRNA targets. *Cell*, 115(7):787–798, 2003. 19, 25, 40, 99

[86] Yanli Wang, Stefan Juranek, Haitao Li, Gang Sheng, Thomas Tuschl, and Dinshaw J. Patel. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, 456(7224):921–926, 2008. 19

[87] Jean Hausser and Mihaela Zavolan. Identification and consequences of miRNA-target interactions–beyond repression of gene expression. *Nature Reviews Genetics*, 15(9):599–612, 2014. 19

[88] D. P. Bartel. MicroRNAs: Target recognition and regulatory functions. *Cell*, 136(2):215–233, 2009. 19, 99

[89] Benjamin P. Lewis, Christopher B. Burge, and David P. Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20, 2005. 19, 20, 25

[90] Chanseok Shin, Jin-Wu Nam, Kyle Kai-How Farh, H. Rosaria Chiang, Alena Shkumatava, and David P. Bartel. Expanding the microRNA targeting code: Functional sites with centered pairing. *Molecular Cell*, 38(6):789–802, 2010. 19, 25

[91] Azra Krek, Dominic Grün, Matthew N. Poy, Rachel Wolf, Lauren Rosenberg *et al.* Combinatorial microRNA target predictions. *Nature genetics*, 37(5):495–500, 2005. 20, 27, 99

[92] Praveen Sethupathy, Molly Megraw, and Artemis G. Hatzigeorgiou. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature Methods*, 3(11):881–886, 2006. 20

[93] Nikolaus Rajewsky. microRNA target predictions in animals. *Nature genetics*, 38 Suppl(6s):S8–S13, 2006. 20, 99

[94] H. Robins, Y. Li, and R. W. Padgett. Incorporating structure to predict microRNA targets. *Proceedings of the National Academy of Sciences of the United States of America*, 102(11):4006–4009, 2005. 21

[95] Ivo L. Hofacker. How microRNAs choose their targets. *Nature Genetics*, 39(10):1191–1192, 2007. 21

[96] Johnson,E., Srivastava,R. Volatility in mRNA secondary structure as a design principle for antisense. *Nucleic Acids Research*, 41(3):e43, 2013. 21, 41, 76, 77

[97] Long,D. *et al.* (2007) Potent effect of target structure on microRNA function. *Nature Structural & Molecular Biology*, 14:287–294, 2007. 21, 28, 77, 103, 110

[98] Ye Ding, Chi Yu Chan, and Charles E. Lawrence. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Research*, 32(suppl 2):W135–W141, 2004. 21, 28

[99] Ray M. Marìn and Jirì Vanìcek. Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Research*, 39(1):19–29, 2011. 21, 66

[100] Stephan H. Bernhart, Ivo L. Hofacker, and Peter F. Stadler. Local RNA base pairing probabilities in large sequences. *Bioinformatics*, 22(5):614–615, 2006. 22

[101] Mückstein, Ulrike and Tafer, Hakim and Hackermüller, Jörg and Bernhart, Stephan H. and Stadler, Peter F. and Hofacker, Ivo L. Thermodynamics of RNARNA binding. *Bioinformatics*, 22(10):1177–1182, 2006. 22

[102] Bompfünewerer, Athanasius F. and Backofen, Rolf and Bernhart, Stephan H. and Hertel, Jana and Hofacker, Ivo L. and Stadler, Peter F. and Will, Sebastian. Variations on RNA folding and alignment: lessons from Benasque. *Journal of Mathematical Biology*, 56(1):129–144, 2008. 22

[103] Bernhart, Stephan H. and Mückstein, Ullrike and Hofacker, Ivo L. RNA Accessibility in cubic time. *Algorithms for Molecular Biology*, 6(1):3, 2011. 22

[104] Subkhanhulova,T. *et al.* Modelling and measuring single cell RNA expression levels find considerable transcriptional differences among phenotypically identical cells. *BMC Genomics*, 9(1):268, 2008. 22, 74, 79, 94

[105] Yu Liang, Dana Ridzon, Linda Wong, and Caifu Chen. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, 8(1):166, 2007. 22

[106] C. Ragan, M. Zuker, and M. A. Ragan. Quantitative prediction of miRNA-mRNA interaction based on equilibrium concentrations. *PLOS Computational Biology*, 7(2):1–11, 2011. 22, 76

[107] Garcia,D.M. *et al.* Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lsy-6* and other microRNAs. *Nature Structural & Molecular Biology*, 18:1139–1146, 2011. 22, 25, 68, 75

[108] Stijn van Dongen Cei Abreu-Goodger, and Anton J Enright. Detecting microRNA binding and siRNA off-target effects from expression data. *Algorithms for Molecular Biology*, 5(12):1023–1025, 2008. 22, 30

[109] J. Robin Lytle, Therese A. Yario, and Joan A. Steitz. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5'UTR as in the 3'UTR. *Proceedings of the National Academy of Sciences*, 104(23):9667–9672, 2007. 23

[110] Inhan Lee, Subramanian S. Ajay, Jong In Yook, Hyun Sil Kim, Su Hyung Hong *et al.* New class of microRNA targets containing simultaneous 5'-utr and 3'-utr interaction sites. *Genome Research*, 19(7):1175–1183, 2009. 23

[111] Zhuo Fang and Nikolaus Rajewsky. The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLoS ONE*, 6(3):1–6, 2011. 23

[112] Jean Hausser, Afzal Pasha Syed, Biter Bilen, and Mihaela Zavolan. Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Research*, 23(4):604–615, 2013. 23

[113] Stark, Alexander AND Brennecke, Julius AND Russell, Robert B AND Cohen, Stephen M. Identification of Drosophila MicroRNA Targets. *PLoS Biology*, 1(3):e60, 2003. 23

[114] Dong Yue, Hui Liu, and Yufei Huang. Survey of Computational Algorithms for MicroRNA Target Prediction. *Current genomics*, 10(7):478–492, 2009. 23

[115] Enright,A.J. *et al.* MicroRNA targets in Drosophila. *Genome Biology*, 5(1):R1, 2003. 24, 25, 40, 99

[116] Bino John, Anton J Enright, Alexei Aravin, Thomas Tuschl, Chris Sander, and Debora S Marks. Human microRNA targets. *PLOS Biology*, 2(11):e363, 2004. 24

[117] Doron Betel, Anjali Koppal, Phaedra Agius, Chris Sander, and Christina Leslie. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biology*, 11(8):R90, 2010. 24, 33

[118] Vikram Agarwal, George W Bell, Jin-Wu Nam, and David P Bartel. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4, 2015. 25, 34

[119] Mathieu Blanchette, W. James Kent, Cathy Riemer, Laura Elnitski, Arian F.A. Smit *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research*, 14(4):708–715, 2004. 25

[120] Griffiths-Jones,S. *et al.* miRBase: tools for microRNA genomics. *Nucleic Acids Research*, 36(suppl 1):D154–D158, 2008. 25

[121] Marc Rehmsmeier, Peter Steffen, Matthias Hochsmann, and Robert Giegerich. Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10):1507–1517, 2004. 26, 29

[122] Jan Krüger and Marc Rehmsmeier. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research*, 34(suppl 2):W451–W454, 2006. 26

[123] Gerd Anders, Sebastian D. Mackowiak, Marvin Jens, Jonas Maaskola *et al.* DoRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Research*, 40(D1):D180–D186, 2012. 27

[124] Kai Blin, Christoph Dieterich, Ricardo Wurmus, Nikolaus Rajewsky, Markus Landthaler, and Altuna Akalin. DoRiNA 2.0: upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Research*, 43(D1):D160–D167, 2015. 27

[125] R. Lorenz, S. Bernhart, C. Honer zu Siederdissen, H. Tafer, C. Flamm, P. Stadler, and I. Hofacker. ViennaRNA package 2.0. *Algorithms for Molecular Biology*, 6(1):26, 2011. 27, 102

[126] Chaochun Liu, Bibekanand Mallick, Dang Long, William A. Rennie, Adam Wolenc *et al.* CLIP-based prediction of mammalian microRNA binding sites. *Nucleic Acids Research*, 41(14):e138, 2013. 29

[127] Chaochun Liu, William A Rennie, Bibekanand Mallick, Shaveta Kanoria, Dang Long *et al.* MicroRNA binding sites in C. elegans 3'UTRs. *RNA Biology*, 11(6):693–701, 2014. 29, 34

[128] William Rennie, Chaochun Liu, C. Steven Carmack, Adam Wolenc, Shaveta Kanoria *et al.* STarMir: a web server for prediction of microRNA binding sites. *Nucleic Acids Research*, 42(W1):W114–W118, 2014. 29

[129] Marianthi Kiriakidou, Peter T. Nelson, Andrei Kouranov, Petko Fitziev, Costas Bouyioukos *et al.*A combined computational-experimental approach predicts human microRNA targets. *Genes & Development*, 18(10):1165–1178, 2004. 30

[130] Manolis Maragkakis, Panagiotis Alexiou, Giorgio Papadopoulos, Martin Reczko, Theodore Dalamagas *et al.* Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, 10(1):295, 2009. 30

[131] Martin Reczko, Manolis Maragkakis, Panagiotis Alexiou, Ivo Grosse, and Artemis G. Hatzigeorgiou. Functional microRNA targets in protein coding sequences. *Bioinformatics*, 28(6):771–776, 2012. 30, 34

[132] Maria D. Paraskevopoulou, Georgios Georgakilas, Nikos Kostoulas, Ioannis S. Vlachos, Thanasis Vergoulis *et al.* DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Research*, 41(W1):W169–W173, 2013. 30

[133] Bartonicek, Nenad and Enright, Anton J. SylArray: a web server for automated detection of miRNA effects from expression data. *Bioinformatics*, 26(22):2900–2901, 2010. 30, 31

[134] Ioannis S. Vlachos, Maria D. Paraskevopoulou, Dimitra Karagkouni, Georgios Georgakilas, Thanasis Vergoulis *et al.* DIANA-Tarbase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Research*, 43(D1):D153–D159, 2015. 31

[135] Feifei Xiao, Zhixiang Zuo, Guoshuai Cai, Shuli Kang, Xiaolian Gao, and Tongbin Li. miRecords: an integrated resource for microRNAtarget interactions. *Nucleic Acids Research*, 37(suppl 1):D105–D110, 2009. 31, 35

[136] Sheng-Da Hsu, Yu-Ting Tseng, Sirjana Shrestha, Yu-Ling Lin, Anas Khaleel *et al.* miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Research*, 42(D1):D78–D85, 2014. 31, 40, 42

[137] Jian-Hua Yang, Jun-Hao Li, Peng Shao, Hui Zhou, Yue-Qin Chen, and Liang-Hu Qu. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Research*, 39(suppl 1):D202–D209, 2011. 31

[138] Jun-Hao Li, Shun Liu, Hui Zhou, Liang-Hu Qu, Jian-Hua Yang starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and proteinRNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, 42(D1):D92–D97, 2014. 31, 64

[139] Kim D. Pruitt, Tatiana Tatusova, and Donna R. Maglott. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 35(suppl 1):D61–D65, 2007. 32

[140] Fiona Cunningham, M. Ridwan Amode, Daniel Barrell, Kathryn Beal, Konstantinos Billis *et al.* Ensembl 2015. *Nucleic Acids Research*, 43(D1):D662–D669, 2015. 32

[141] Kate R. Rosenbloom, Joel Armstrong, Galt P. Barber, Jonathan Casper, Hiram Clawson *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, 43(D1):D670–D681, 2015. 32

[142] Shanrong Zhao and Baohong Zhang. A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genomics*, 16(1):97, 2015. 32

[143] Doron Betel, Manda Wilson, Aaron Gabow, Debora S. Marks, and Chris Sander. The microRNA.org resource: targets and expression. *Nucleic Acids Research*, 36(suppl 1):D149–D153, 2008. 32

[144] Robert Petryszak, Tony Burdett, Benedetto Fiorelli, Nuno A. Fonseca, Mar Gonzalez-Porta *et al.* Expression Atlas update - a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Research*, 42(D1):D926–D932, 2014. 32

[145] Kevin C. Miranda, Tien Huynh, Yvonne Tay, Yen-Sin Ang, Wai-Leong Tam *et al.* A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, 126(6):1203–1217, 2006. 34

[146] Amit Kumar, Adam K.-L. Wong, Mark L. Tizard, Robert J. Moore, and Christophe Lefevre. miRNA_ Targets: A database for miRNA target predictions in coding and non-coding regions of mRNAs. *Genomics*, 100(6):352–356, 2012. 34

[147] Xiao Fan and Lukasz Kurgan. Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Briefings in Bioinformatics*, 16(5):780–794, 2015. 34

[148] Charles E. Vejnar and Evgeny M. Zdobnov. MiRmap: Comprehensive prediction of microRNA target repression strength. *Nucleic Acids Research*, 40(22):11673–11683, 2012. 34

[149] Nathan Wong and Xiaowei Wang. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Research*, 43(D1):D146–D152, 2015. 34

[150] Seunghak Yu, Juho Kim, Hyeyoung Min, and Sungroh Yoon. Ensemble learning can significantly improve human microRNA target prediction. *Methods*, 69(3):220–229, 2014. 35

[151] Neil H. Tan Gana, Ann F. B. Victoriano, and Takashi Okamoto. Evaluation of online miRNA resources for biomedical applications. *Genes to Cells*, 17(1):11–27, 2012. 35

[152] Bush, William S. AND Moore, Jason H. Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology*, 8(12):1–11, 2012. 35, 36

[153] Cariaso, Michael and Lennon, Greg. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Research*, 40(D1):D1308–D1312, 2012. 35

[154] Julienne M. Mullaney, Ryan E. Mills, W. Stephen Pittard, and Scott E. Devine. Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics*, 19(R2):R131–R136, 2010. 35

[155] David W. Salzman and Joanne B. Weidhaas. SNPing cancer in the bud: microRNA and microRNA-target site polymorphisms as diagnostic and prognostic biomarkers in cancer. *Pharmacology & Therapeutics*, 137(1):55–63, 2013. 35, 36

[156] Xu Zhao, Qing Ye, Kang Xu, Jinluo Cheng, Yanqin Gao *et al.* Single-nucleotide polymorphisms inside microRNA target sites influence the susceptibility to type 2 diabetes. *Journal of Human Genetics*, 58:135–141, 2013. 36

[157] Jana Petrkova, Amit Kishore, Jana Borucka and Martin Petrek. Novel Insights into miRNA in Lung and Heart Inflammatory Diseases. *Mediators of Inflammation*, 2014, 2014. 36

[158] Gaofeng Wang, Joelle M. van der Walt, Gregory Mayhew, Yi-Ju Li, Stephan Zuechner *et al.* Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. *American Journal of Human Genetics*, 82(2):283–289, 2008. 36, 47

[159] S. T. Sherry, M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, 29(1):308–311, 2001. 36

[160] Peter M. Visscher and Matthew A. Brown and Mark I. McCarthy and Jian Yang. Five Years of {GWAS} Discovery. *The American Journal of Human Genetics*, 90(1):7–24, 2012. 36

[161] Calin, George Adrian and Dumitru, Calin Dan and Shimizu, Masayoshi and Bichi, Roberta *et al.* Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences*, 99(24):15524–15529, 2002. 36

[162] Calin , George Adrian and Ferracin , Manuela and Cimmino , Amelia *et al.* A MicroRNA Signature Associated with Prognosis and Progression in Chronic Lymphocytic Leukemia. *New England Journal of Medicine*, 353(17):1793–1801, 2005. 36

[163] Abelson, Jesse F. and Kwan, Kenneth Y. and O'Roak, Brian J. and Baek, Danielle Y. *et al.* Sequence Variants in SLITRK1 Are Associated with Tourette's Syndrome. *Science*, 310(5746):317–320, 2005. 36

[164] Clop, Alex and Marcq, Fabienne and Takeda, Haruko and Pirottin, Dimitri *et al.* A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature genetics*, 38(7):813–818, 2006. 36, 37

[165] Chin, Lena J. and Ratner, Elena and Leng, Shuguang and Zhai, Rihong *et al.* A SNP in a let-7 microRNA Complementary Site in the KRAS 3' Untranslated Region Increases Non–Small Cell Lung Cancer Risk. *Cancer Research*, 68(20):8535–8540, 2008. 37

[166] Saunders, M. A. and Liang, H. and Li, W. H. Human polymorphism at microRNAs and microRNA target sites. *Proceedings of the National Academy of Sciences*, 104(9):3300–3305, 2007. 37

[167] Zorc, Minja AND Jevsinek Skok, Dasa AND Godnic, Irena AND Calin, George Adrian *et al.* Catalog of MicroRNA Seed Polymorphisms in Vertebrates. *PLOS ONE*, 7(1):1–8, 2012. 37

[168] Mencia, Angeles and Modamio-Hoybjor, Silvia and Redshaw, Nick and Morin, Matias *et al.* Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nature Genetics*, 41(5):609–613, 2009. 37

[169] Sayed, Danish and Abdellatif, Maha. MicroRNAs in Development and Disease. *Physiological Reviews*, 91(3):827–887, 2011. 37

[170] Bruno,A. *et al.* miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3'UTRs of human genes. *BMC Bioinformatics*, 13(1):44, 2012. 37, 42

[171] Anindya Bhattacharya, Jesse D. Ziebarth, and Yan Cui. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Research*, 42(D1):D86–D91, 2014. 37

[172] Barenboim,M. *et al.* MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. *Human Mutation*, 31(11):1223–1232, 2010. 37

[173] Hiard, Samuel and Charlier, Carole and Coppieters, Wouter and Georges, Michel and Baurain, Denis. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Research*, 38(suppl 1):D640–D651, 2010. 37

[174] Praveen Sethupathy and Francis S. Collins. MicroRNA target site polymorphisms and human disease. *Trends in Genetics*, 24(10):489–497, 2008. 37, 43, 97

[175] T. Vergoulis, I. S. Vlachos, P. Alexiou, G. Georgakilas, M. Maragkakis, M. Reczko, S. Gerangelos, N. Koziris, T. Dalamagas, and A. G. Hatzigeorgiou. Tarbase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Research*, 40(D1):D222–D229, 2011. 40

[176] E. Wienholds and R. H. Plasterk. Microrna function in animal development. *FEBS Letters*, 579(26):5911–5922, 2005. 40

[177] Iorio, Marilena V., and Carlo M. Croce. Causes and Consequences of microRNA Dysregulation. *Cancer journal*, 18(3):215-222, 2012. 40

[178] Johnson AD, Trumbower H, Sadee W. RNA Structures Affected By Single Nucleotide Polymorphisms In Transcribed Regions Of The Human Genome. *WebmedCentral BIOINFORMATICS*, 2(2):WMC001600, 2011. 40, 41, 117

[179] Martin Joshua S., Halvorsen Matthew, Davis-Neulander Lauren, Ritz Justin, Gopinath Chetna *et al.* Structural effects of linkage disequilibrium on the transcriptome. *RNA*, 18(1):77–87, 2012. 41, 117

[180] Ulrike Haas, Georg Sczakiel, and Sandra Laufer. MicroRNA-mediated regulation of gene expression is affected by disease-associated SNPs within the 3'-UTR via altered RNA structure. *RNA Biology*, 9(6):924–937, 2012. 41, 47

[181] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao, and Q. Cui. An analysis of human microRNA and disease associations. *PLoS ONE*, 3(10):1–5, 2008. 41

[182] Teo,M. *et al.* The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes. *Carcinogenesis*, 33(3):581–586, 2012. 43, 46, 48, 53, 87

[183] S Garritano, C De Santi, R Silvestri, O Melaiu, M Cipollini*et al.* A common polymorphism within MSLN affects miR-611 binding site and soluble

mesothelin levels in healthy people. *Journal of Thoracic Oncology*, 9(11): 1662–1668, 2014. 43

[184] Hao Zhang, Yunqiang Liu, Dan Su, Yuan Yang, Gang Bai *et al.* A single nucleotide polymorphism in a miR-1302 binding site in CGA increases the risk of idiopathic male infertility. *Fertility and Sterility*, 96(1):34–39, 2011. 43

[185] Kalabus,J.L. *et al.* MicroRNAs differentially regulate carbonyl reductase 1 (CBR1) gene expression dependent on the allele status of the common polymorphic variant rs9024. *PLoS One*, 7(11):1–9 , 2012. 43

[186] Kapeller Johannes, Houghton Lesley A., Mönnikes Hubert, Walstab Jutta, Möller Dorothee *et al.* First evidence for an association of a functional variant in the microRNA-510 target site of the serotonin receptor-type 3E gene with diarrhea predominant irritable bowel syndrome. *Human Molecular Genetics*, 17(19):2967–2977, 2008. 43

[187] S. Y. Lee, J. E. Choi, H. S. Jeon, M. J. Hong, Y. Y. Choi *et al.* A genetic variation in microRNA target site of KRT81 gene is associated with survival in early-stage non-small-cell lung cancer. *Annals of Oncology*, 26(6):1142–1148, 2015. 43

[188] K. Hikami, A. Kawasaki, I. Ito, M. Koga, S. Ito *et al.* Association of a functional polymorphism in the 3'-untranslated region of SPI1 with systemic lupus erythematosus. *Arthritis & Rheumatism*, 63(3):755–763, 2011. 44

[189] Tan,Z. *et al.* Allele-specific targeting of microRNAs to HLA-G and risk of asthma. *The American Society of Human Genetics*, 81(4):829–834, 2007. 44, 87

[190] S. Minguzzi, S. D. Selcuklu, C. Spillane, and A. Parle-McDermott. An NTD-associated polymorphism in the 3'UTR of MTHFD1L can affect disease risk by altering miRNA binding. *Human Mutation*, 35(1):96–104, 2014. 44

[191] Shunxin Song, Dianke Chen, Jiachun Lu, Jiawei Liao, Yanxin Luo *et al.* NFkB1 and NFkBIA polymorphisms are associated with increased risk for sporadic colorectal cancer in a southern chinese population. *PloS one*, 6(6): 1–10, 2011. 44, 53

[192] Jiali Xu, Shengwang Tian, Zhiqiang Yin, Shuangshuang Wu, Lingxiang Liu *et al.* MicroRNA-binding site SNPs in deregulated genes are associated with clinical outcome of non-small cell lung cancer. *Lung Cancer*, 85(3): 442–448, 2014. 44

[193] Ying Liu, Hongbing Cai, Jie Liu, Haoning Fan, Zhiyuan Wang *et al.* A miR-151 binding site polymorphism in the 3'-untranslated region of the cyclin E1 gene associated with nasopharyngeal carcinoma. *Biochemical and Biophysical Research Communications*, 432(4):660–665, 2013. 45

[194] Chi Keung Cheng, Tsz Ki Kwan, Chi Ying Cheung, Kitty Ng, Pei Liang *et al.* A polymorphism in the 3'-untranslated region of the NPM1 gene causes illegitimate regulation by microRNA-337-5p and correlates with adverse outcome in acute myeloid leukemia. *Haematologica*, 98(6):913–917, 2013. 45

[195] Brahmaraju Mopidevi, Madhusudhan Ponnala, and Ashok Kumar. Human angiotensinogen +11525 C/A polymorphism modulates its gene expression through microRNA binding. *Physiological Genomics*, 45(19):901–906, 2013. 45

[196] Charlotte Delay, Veronique Dorval, Alice Fok, Benjamin Grenier-Boley, Jean-Charles Lambert *et al.* MicroRNAs targeting Nicastrin regulate Aβ production and are affected by target site polymorphisms. *Frontiers in Molecular Neuroscience*, 7:67, 2014. 45

[197] Barbara Pardini, Fabio Rosa, Elisa Barone, Cornelia Di Gaetano, Jana Slyskova *et al.* Variation within 3'UTRs of base excision repair genes and response to therapy in colorectal cancer patients: A potential modulation of microRNAs binding. *Clinical Cancer Research*, 19(21):6044–6056, 2013. 45

[198] Cipollini, Monica, Stefano Landi, and Federica Gemignani. MicroRNA Binding Site Polymorphisms as Biomarkers in Cancer Management and Research. *Pharmacogenomics and Personalized Medicine*, 7:173–191, 2014. 46

[199] Reka Kovacs-Nagy, Zsuzsanna Elek, Anna Szekely, Tibor Nanasi, Maria Sasvari-Szekely, and Zsolt Ronai. Association of aggression with a novel microRNA binding site polymorphism in the wolframin gene. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 162(4): 404–412, 2013. 46

[200] L. Ma, M. Xu, D. Li, Y. Han, Z. Wang *et al.* A miRNA-binding-site SNP of MSX1 is associated with NSOC susceptibility. *Journal of Dental Research*, 93(6):559–564, 2014. 46

[201] Y. Li, Y. Nie, J. Cao, S. Tu, Y. Lin, Y. Du, and Y. Li. G-A variant in miR-200c binding site of EFNA1 alters susceptibility to gastric cancer. *Molecular Carcinogenesis*, 53(3):219–229, 2014. 46

[202] Zwiers,A. *et al.* Cutting Edge: A Variant of the IL-23R Gene Associated with Inflammatory Bowel Disease Induces Loss of MicroRNA Regulation and Enhanced Protein Production. *The Journal of Immunology*, 188(4):1573–1577, 2012. 46, 87

[203] Wang,L. *et al.*, (2012) A miRNA binding site single-nucleotide polymorphism in the 3'-UTR region of the IL23R gene is associated with breast cancer. *PLoS One*, 7(12):1–6, 2012. 47

[204] Zhang,L. *et al.* Functional SNP in the microRNA-367 binding site in the 3'UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. *Proceedings of the National Academy of Sciences*, 108(33):13653–13658, 2011. 47

[205] Lin Yuan, Haiyan Chu, Meilin Wang, Xiaojian Gu, Danni Shi *et al.* Genetic variation in *D*ROSHA 3'UTR regulated by hsa-miR-27b is associated with bladder cancer risk. *PLoS ONE*, 8(11):e81524, 2013. 47

[206] Luo,J. *et al.* A microRNA-7 binding site polymorphism in HOXB5 leads to differential gene expression in bladder cancer. *PLoS One*, 7(6):1–10, 2012 47

[207] S. Zhang, H. Chen, X. Zhao, J. Cao, J. Tong, J. Lu, W. Wu, H. Shen, Q. Wei, and D. Lu. REV3L 3'UTR 460 T-C polymorphism in microRNA target sites contributes to lung cancer susceptibility. *Oncogene*, 32:242-250;, 2012. 48

[208] Chang,W. *et al.* ORAI1 genetic polymorphisms associated with the susceptibility of atopic dermatitis in Japanese and Taiwanese populations. *PLoS One*, 7(1):1–8, 2012. 48, 94, 97

[209] Wang,K. *et al.* MiR-196a binding-site SNP regulates RAP1A expression contributing to esophageal squamous cell carcinoma risk and metastasis. *Carcinogenesis*, 33(11):2147–2154, 2012. 48

[210] C. Delay, F. Calon, P. Mathews, and S. Hebert. Alzheimer-specific variants in the 3'UTR of Amyloid precursor protein affect microRNA function. *Molecular Neurodegeneration*, 6(1):70, 2011. 48

[211] Yuzhen Gao, Yan He, Jun Ding, Kang Wu, Bo Hu, Yang Liu, Yan Wu, Bingnan Guo, Yueping Shen, Debora Landi, Stefano Landi, Yifeng Zhou, and Haiyan Liu. An insertion/deletion polymorphism at miRNA-122-binding site in the interleukin-1alpha 3' untranslated region confers risk for hepatocellular carcinoma. *Carcinogenesis*, 30(12):2064–2069, 2009. 49

[212] M. Cheng, L. Yang, R. Yang, X. Yang, J. Deng *et al.* A microRNA-135a/b binding polymorphism in CD133 confers decreased risk and favorable prognosis of lung cancer in Chinese by reducing CD133 expression. *Carcinogenesis*, 34(10):2292–2299, 2013. 49

[213] Guoyu Zhang, Na Li, Zhu Li, Qianqian Zhu, Fang Li *et al.* MicroRNA-4717 differentially interacts with its polymorphic target in the PD1 3' untranslated region: A mechanism for regulating PD-1 expression and function in HBV-associated liver diseases. *Oncotarget*, 6(22):18933–44, 2015. 49

[214] Xu, Zongli and Taylor, Jack A. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Research*, 37(suppl 2):W600–W605, 2009. 53

[215] Liu Chenxing, Zhang Fuquan, Li Tingting, Lu Ming, Wang Lifang *et al.* MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics*, 13(1):661, 2012. 53

[216] Deveci, Mehmet and Çatalyürek, Ümit V. and Toland, Amanda Ewart mrSNP: Software to detect SNP effects on microRNA binding. *BMC Bioinformatics*, 15(1):1–10, 2014. 56

[217] Thomas, Laurent F. and Saito, Takaya and Sætrom, P. Inferring causative variants in microRNA target sites. *Nucleic Acids Research*, 39(16):e109, 2011. 56

[218] Gong, Jing and Liu, Chunjie and Liu, Wei and Wu, Yuliang and Ma, Zhaowu and Chen, Hu and Guo, An-Yuan An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. *Database*, 2015:bav029, 2015. 57

[219] Dweep, H *et al.* miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nature Methods*, 12(8):697–697, 2015. 60

[220] Zhang Y, Verbeek FJ Comparison and integration of target prediction algorithms for microRNA studies. *Journal of Integrative Bioinformatics*, 7(3):127, 2010. 66

[221] L. Day, O. Abdelhadi Ep Souki, A. A. Albrecht, and K. Steinhöfel. Accessibility of microRNA binding sites in metastable RNA secondary structures in the presence of SNPs. *Bioinformatics*, 30(3):343–352, 2014. 74

[222] Arvey,A *et al.* Target mRNA abundance dilutes microRNA and siRNA activity. *Molecular Systems Biology*, 6(1):363, 2010. 74, 75

[223] Salmena,L. *et al.* A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? *Cell*, 146(3):353–358, 2011. 75

[224] Saito,T., Sætrom,P. Target gene expression levels and competition between transfected and endogenous microRNAs are strong confounding factors in microRNA high-throughput experiments. *Silence*, 3(1):3, 2012. 75

[225] Mullokandov,G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nature Methods*, 9:840–846, 2012. 75

[226] Larsson,E. *et al.* mRNA turnover rate limits siRNA and microRNA efficacy. *Molecular Systems Biology*, 6(1):433, 2010. 75

[227] Cuccato,G. *et al.* Modeling RNA interference in mammalian cells. *BMC Systems Biology*, 5(1):19, 2011. 75, 79

[228] Osella,M. *et al.* The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLOS Computational Biology*, 7(3):1–16, 2011. 75

[229] Loinger,A. *et al.* Competition between small RNAs: A quantitative view. *Biophysical Journal*, 102(8):1712–1721, 2012. 76

[230] Baker,C. *et al.* Stochastic modeling of regulation of gene expression by multiple small RNAs. *Physical Review E*, 85(6):061915, 2012. 76

[231] Marin,R.M., Vaniček,J. Optimal use of conservation and accessibility filters in microRNA target prediction. *PLoS One*, 7(2):1–11, 2012. 76

[232] Muniategui,A. *et al.* Joint analysis of miRNA and mRNA expression data. *Briefings in Bioinformatics*, 14(3):263–278, 2013. 76, 99

[233] Shirdel,E. *et al.* NAViGaTing the micronome - using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PLoS One*, 6(2):1–17, 2011. 76, 99

[234] Gennarino,V. *et al.* Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Research*, 22(6):1163–1172, 2012. 76

[235] Balaga,O. *et al.* Toward a combinatorial nature of microRNA regulation in human cells. *Nucleic Acid Research*, 40(19):9404–9416, 2012. 76

[236] Ding,Y. *et al.* RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA*, 11(8):1157–1166, 2005. 77

[237] Gruber,A.R. *et al.* The Vienna RNA Websuite. *Nucleic Acids Research*, 36(Web Server issue):W70–W74, 2008. 77, 88, 101, 103, 118

[238] Wolfinger,M.T. *et al.* Efficient computation of RNA folding dynamics. *Journal of Physics A: Mathematical and General*, 37(17):4731–4741, 2004. 77

[239] Hammell,M. *et al.* mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nature Methods*, 5:813–819, 2008. 78, 79

[240] Chen,J.L. *et al.* Testing the Nearest Neighbor Model for canonical RNA base pairs: revision of GU parameters. *Biochemistry*, 51(16):3508–3522, 2012. 78, 103

[241] Ye,W. *et al.* The effect of central loops in miRNA:MRE duplexes on the efficiency of miRNA-mediated gene regulation. *PLoS One*, 3(3):1–12, 2008. 81

[242] Abdelhadi Ep Souki, Ouala and Day, Luke and Albrecht, Andreas A. and Steinhöfel, Kathleen MicroRNA Target Prediction Based Upon Metastable RNA Secondary Structures. *Bioinformatics and Biomedical Engineering: Third International Conference, IWBBIO 2015, Granada, Spain, April 15-17, 2015. Proceedings, Part II*, 12(8):697–697, 2015. 99, 109

[243] P. Alexiou, M. Maragkakis, G. L. Papadopoulos, M. Reczko, and A. G. Hatzigeorgiou. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, 25(23):3049–3055, 2009. 99

[244] S. M. Peterson, J. A. Thompson, M. L. Ufkin, P. Sathyanarayana, L. Liaw, and C. B. Congdon. Common features of microRNA target prediction tools. *Frontiers in Genetics*, 5:23, 2014. 99

[245] Zhao, Yong and Samal, Eva and Srivastava, Deepak. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, 436:214–220, 2005. 99

[246] Connerty, Patrick and Ahadi, Alireza and Hutvagner, Gyorgy. RNA Binding Proteins in the miRNA Pathway. *International Journal of Molecular Sciences*, 17(1):31, 2016. 117

[247] Li, Xiao and Quon, Gerald and Lipshitz, Howard D. and Morris, Quaid. Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA*, 16(6):1096–1107, 2010. 117

[248] Li, Xiao and Kazan, Hilal and Lipshitz, Howard D. and Morris, Quaid D. Finding the target sites of RNA-binding proteins. *Wiley Interdisciplinary Reviews: RNA*, 5(1):111–130, 2014. 117

[249] Westra,H. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics* 45:1238-1243, 2013. 117

List of Figures

List of Tables